ICPSR | INTER-UNIVERSITY CONSORTIUM FOR POLITICAL AND SOCIAL RESEARCH

ICPSR 34933

National Survey on Drug Use and Health, 2012

United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Center for Behavioral Health Statistics and Quality

Methodological Resources

Inter-university Consortium for
Political and Social Research
P.O. Box 1248
Ann Arbor, Michigan 48106
www.icpsr.umich.edu

Terms of Use

The terms of use for this study can be found at:
http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies//terms

Information about Copyrighted Content

Some instruments administered as part of this study may contain in whole or substantially in part contents from copyrighted instruments. Reproductions of the instruments are provided as documentation for the analysis of the data associated with this collection. Restrictions on "fair use" apply to all copyrighted content. More information about the reproduction of copyrighted works by educators and librarians is available from the United States Copyright Office.

NOTICE
WARNING CONCERNING COPYRIGHT RESTRICTIONS

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use," that user may be liable for copyright infringement.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.103
Contract No. HHSS283201000003C
Phase I, Deliverable 8

Authors:

Katherine B. Morton
Peilan C. Martin
Bonnie E. Shook-Sa
James R. Chromy
Erica L. Hirsch

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLE DESIGN REPORT

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.103
Contract No. HHSS283201000003C
Phase I, Deliverable No. 8

Authors:

Katherine B. Morton
Peilan C. Martin
Bonnie E. Shook-Sa
James R. Chromy
Erica L. Hirsch

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

January 2013

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. Significant contributors at RTI include Katherine B. Morton, Peilan C. Martin, Bonnie E. Shook-Sa, James R. Chromy, Erica L. Hirsch, Jeremy Aldworth, Patrick Chen, David Chrest, Phillip Kott, Richard Straw, Debbie Bond, Valerie Garner, Teresa Bass, and Thomas G. Virag (Project Director).

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Chapter **Page**

List of Tables v

List of Exhibits vii

1. Overview 1
   - 1.1 Target Population 1
   - 1.2 Design Overview 1
   - 1.3 5-Year Design and the 2010-2011 and 2012-2013 Sample Extensions 2
   - 1.4 Stratification and First- and Second-Stage Sample Selections 2
   - 1.5 Sample Dwelling Units and Persons 4

2. Extending the Coordinated 5-Year Sample 7
   - 2.1 Formation of and Objectives for Using the Composite Size Measures 7
   - 2.2 Stratification 8
   - 2.3 First- and Second-Stage Sample Selection 9
   - 2.4 Survey Year and Quarter Assignment 11
   - 2.5 Creation of Variance Estimation Strata and Replicates 11
   - 2.6 Other Sampling-Related Variables 13

3. General Sample Allocation Procedures for the Main Study 15
   - 3.1 Notation 16
   - 3.2 Determining Person Sample Sizes by State and Age Group 18
   - 3.3 Third-Stage Sample Allocation for Each Segment 22
     - 3.3.1 Dwelling Unit Frame Construction—Counting and Listing 22
     - 3.3.2 Determining Dwelling Unit Sample Size 26
   - 3.4 Determining Fourth-Stage Sample (Person) Selection Probabilities for Each Segment 26
   - 3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden 26
   - 3.6 Dwelling Unit Selection and Release Partitioning 27
   - 3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units 28
   - 3.8 Quarter-by-Quarter Deviations 33
   - 3.9 Sample Weighting Procedures 35
     - 3.9.1 Main Study Sampling Weights 35
     - 3.9.2 Quality Control Measures in Design-Based Weighting Procedures 43

4. General Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study 45
   - 4.1 Respondent Universe and Sampling Methods for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study 45
   - 4.2 Person Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study 46

Table of Contents (continued)

Chapter **Page**

5. General Sample Procedures for the Questionnaire Field Test 51
5.1 Target Population 51
5.2 Selection of State Sampling Regions and Segments 51
5.3 Selection of Dwelling Units 53
5.4 Age Group Allocations 54
5.5 Selection of Persons 54
5.6 Creation of Variance Estimation Strata and Replicates 55

References 57

Appendix

A 2005 through 2013 NSDUH State Sampling Regions 59
B 2012 NSDUH Procedure for Subsegmenting 111
C 2012 NSDUH Procedure for Adding Missed Dwelling Units 113

List of Tables

Table **Page**

2.1 Number of Census Tracts and Segments on Sampling Frame, by State 10
2.2 Segment Identification Number Suffixes and Quarter Assignment 12
3.1 Expected Relative Standard Errors, by Age Group 20
3.2 Expected Main Study Sample Sizes, by State and Age Group 20
3.3 Number of Map Pages, by State and Segment 23
3.4 Segment and Dwelling Unit Summary 24
3.5 Quarterly Sample Sizes and Percentages Released 29
3.6 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment 37
3.7 Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level 38
3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment 39
3.9 Model Group Definitions (Census Divisions) 40
4.1 MHSS/EMHSS Age-Related Factors 46
4.2 Projected Yields of Predicted Positive Cases 47
4.3 2012 MHSS/EMHSS Sample Allocation, by K6 Group 48
4.4 2012 MHSS/EMHSS Sample Allocation, by WHODAS Score 48
5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State 52
5.2 Summary of the 2012 Questionnaire Field Test Sample Results 54

List of Exhibits

Exhibit **Page**

1 Flowchart of Sample Weighting Steps 41
2 Sample Weight Components.....42
3 MHSS/EMHSS Sample Weight Components 49

1. Overview

1.1 Target Population

The respondent universe for the 2012 National Survey on Drug Use and Health[1] (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States and the District of Columbia. Consistent with the NSDUH designs since 1991, the 2012 NSDUH universe included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Coverage before the 1991 survey was limited to residents of the coterminous 48 States, and it excluded residents of group quarters and all persons (including civilians) living on military bases. Persons excluded from the 2012 universe included those with no fixed household address (e.g., homeless and/or transient persons not in shelters) and residents of institutional group quarters, such as jails and hospitals.

1.2 Design Overview

Beginning in 1999 and continuing through subsequent years, the Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the way NSDUH would be conducted. The surveys are conducted using computer-assisted interviewing (CAI) methods and provide improved State estimates based on minimum sample sizes per State. The national sample size of 67,500 is equally allocated across three age groups: persons aged 12 to 17, persons aged 18 to 25, and persons aged 26 or older. This large sample size allows SAMHSA to continue reporting precise demographic subgroups at the national level without needing to oversample specially targeted demographics, as required in the past. This large sample is referred to as the "main sample." The achieved sample for the 2012 NSDUH was 68,309 persons.

Beginning with the 2002 NSDUH and continuing through the 2012 NSDUH, survey respondents were given a $30 incentive payment for participation. As expected, the incentive had the effect of increasing response rates, thereby requiring fewer selected households than previous surveys. In recent years, however, response rates have been slowly declining, which has required the number of selected households to increase.

An additional design change was made in 2002 and continued through 2012. A new pair-sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons on the roster (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons.

A Mental Health Surveillance Study (MHSS) was conducted to monitor the efficacy of a new mental health screening measure by conducting follow-up telephone psychiatric interviews using the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) with selected respondents (First, Spitzer, Gibbon, & Williams,

[1] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

2002).[2] The MHSS (which began in 2008) was conducted as an embedded study within the 2012 NSDUH. The sample design for the MHSS is described in Chapter 4.

Finally, a Questionnaire Field Test (QFT) was conducted in late 2012 to test a redesigned questionnaire protocol for implementation in the 2015 NSDUH. Chapter 5 describes the sample design for the 2012 QFT.

1.3 5-Year Design and the 2010-2011 and 2012-2013 Sample Extensions

A coordinated sample design was developed for the 2005 through 2009 NSDUHs. The 2010-2011 and 2012-2013 samples are extensions of the 5-year sample. Although there is no planned overlap with the 1999 through 2004 samples, the coordinated design for 2005 through 2009 facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. This design was intended to increase the precision of estimates in year-to-year trend analyses, using the expected positive correlation resulting from the overlapping sample between successive NSDUH years. The 2012 NSDUH main sample continues the 50 percent overlap by retaining half of the second-stage units from the 2011 survey. The 2012 design provides for estimates by State in all 50 States plus the District of Columbia. States may therefore be viewed as the first level of stratification and as a reporting variable. Eight States, referred to as the "large" States,[3] had samples designed to yield 3,600 respondents per State for the 2012 main study. This sample size was considered adequate to support direct State estimates. The remaining 43 States[4] had samples designed to yield 900 respondents per State in the 2012 main study. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation (SAE) methodology. Reliable direct State estimates are also possible (in any State) by pooling multiple years of data.

1.4 Stratification and First- and Second-Stage Sample Selections

Within each State, State sampling (SS) regions were formed. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. In other words, regions were formed such that each area yielded, in expectation, roughly the same number of interviews during each data collection period. The smaller States were partitioned into 12 SS regions, whereas the 8 large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions. Maps for these regions, which will be used through the 2013 NSDUH, can be found in Appendix A.

Unlike the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs, the first stage of selection for the 2005 through 2012 NSDUHs was census tracts.[5] This stage was

---

[2] "DSM-IV-TR" stands for the *Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision* (American Psychiatric Association [APA], 2000).

[3] The large States are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

[4] For reporting and stratification purposes, the District of Columbia is treated the same as a State, and no distinction is made in the discussion.

[5] A census tract is a small, relatively permanent statistical subdivision of a county or equivalent entity that contains between 1,200 and 8,000 people, with an optimum size of 4,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

included to contain sample segments within a single census tract to the extent possible.[6] In prior years, segments that crossed census tract boundaries made merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated within SS regions until each tract[7] had, at a minimum, 150 DUs[8] in urban areas and 100 DUs in rural areas.[9]

Before selecting census tracts, additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES[10] (core-based statistical area/socioeconomic status) indicator[11] and by the percentage of the population that is non-Hispanic and white.[12] From this well-ordered sample frame, 48 census tracts per SS region were sequentially selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979).

Because census tracts generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census tract. For this second stage of sampling, each selected census tract was partitioned into compact clusters[13] of DUs by aggregating adjacent census blocks.[14] Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." A sample DU in NSDUH refers to either a housing unit or a group quarters listing unit, such as a dormitory room

---

[6] Some census tracts had to be aggregated in order to meet the minimum DU requirement.

[7] For the remainder of the discussion, first-stage sampling units are referred to as "census tracts" even though each first-stage sampling unit contains one or more census tracts.

[8] DU counts were obtained from the 2000 census data supplemented with revised population counts from Nielsen Claritas.

[9] The basis for the differing minimum DU requirement in urban and rural areas is that it is more difficult to meet the requirement in rural areas, and 100 DUs are sufficient to support one field test and two main study samples.

[10] CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2003).

[11] Four categories are defined as (1) CBSA/low SES, (2) CBSA/high SES, (3) non-CBSA/low SES, and (4) non-CBSA/high SES. To define SES, census tract–level median rents and property values obtained from the 2000 Census Summary File 3 were given a rank (1,…,5) based on State and CBSA quintiles. The rent and value ranks then were averaged, weighted by the percentages of renter- and owner-occupied DUs, respectively. If the resulting score fell in the lower 25th percentile by State and CBSA, the area was considered "low SES"; otherwise, it was considered "high SES."

[12] Although the large sample size eliminates the need for the oversampling of specially targeted demographic subgroups as was required prior to the 1999 NHSDA, sorting by a CBSA/SES indicator and by the percentage of the population that is non-Hispanic and white ensures dispersion of the sample with respect to SES and race/ethnicity. Implicit stratification also has the potential to lower sampling error by reducing the selection of neighboring and possibly similar segments than if the selection was done completely at random.

[13] Although the entire cluster is compact, the final sample of DUs represents a noncompact cluster. Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. Although compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish, 1965).

[14] A census block is a small statistical area bounded by visible features (streets, roads, streams, railroad tracks, etc.) and nonvisible boundaries (e.g., city, town, and county limits). A block group is a cluster of census blocks within the same census tract and generally contains between 300 and 6,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement will support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

Prior to selection, the segments were sorted in the order they were formed (i.e., geographically), and one segment was selected within each sampled census tract using Chromy's method of sequential random sampling (with probability proportionate to size and minimum replacement) (Chromy, 1979). The 48 selected segments then were randomly assigned to a survey year and quarter of data collection as described in Section 2.4.

1.5 Sample Dwelling Units and Persons

After sample segments for the 2012 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the third stage of sample selection.

The primary objective of the third stage of sample selection (listing units) was to determine the minimum number of DUs needed in each segment to meet the targeted sample sizes for all age groups. Thus, listing unit sample sizes for the segment were determined using the age group with the largest sampling rate, which is referred to as the "driving" age group. Using 2000 census data adjusted to more recent data from Claritas, State- and age-specific sampling rates were computed. These rates then were adjusted by the segment's probability of selection; the subsegmentation inflation factor,[15] if any; the probability of selecting a person in the age group (equal to the maximum, or 0.99, for the driving age group); and an adjustment for the "maximum of two" rule.[16] In addition to these factors, historical data from the 2010, 2011, and 2012 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate then was multiplied by the actual number of DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

Some constraints were put on the listing unit sample sizes. For example, to ensure adequate samples for supplemental studies, the listing unit sample size could not exceed 100 per segment or half of the actual listing unit count. Similarly, if five unused listing units remained in the segment, a minimum of five listing units per segment was required for cost efficiency.

Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous

[15] Segments found to be very large in the field are partitioned into *subsegments*. Then one subsegment is chosen at random with probability proportional to the size to be fielded. In some cases, a second-level subsegmenting was required if the census totals used in the initial subsegmenting were off and the selected subsegment was still too large for listing. The subsegmentation inflation factor accounts for reducing the size of the segment.

[16] Brewer's Selection Algorithm never allows for greater than two persons per household to be chosen. Thus, sampling rates are adjusted to satisfy this constraint.

years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, the new or missed dwellings were selected into the 2012 NSDUH using the half-open interval selection technique.[17] This selection technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities, and it also eliminates any bias that might be associated with using "old" segment listings.

Using the roster information obtained from an eligible member of the selected DU, 0, 1, or 2 persons were selected for the survey. Sampling rates were preset by age group and State. Roster information was entered directly into the electronic screening instrument, which automatically implemented this fourth stage of selection based on the State and age group sampling parameters.

One advantage of using an electronic screening instrument in NSDUH is the ability to impose a more complicated person-level selection algorithm on the fourth stage of the NSDUH design. Similar to the 1999 through 2011 designs, one feature that was included in the 2012 design was that any two survey-eligible persons within a DU had some chance of being selected (i.e., all survey-eligible pairs of persons had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of another family member residing in the same DU (e.g., the relationship of drug use between a parent and his or her child).

[17] In summary, this technique states that, if a DU is selected for the 2012 study and an interviewer observes any new or missed DUs between the selected DU and the DU appearing immediately after the selection on the counting and listing form, all new or missed dwellings falling in this interval will be selected. These added DUs are assigned the same probability of selection as the selected DU. If a large number of new or missed DUs are encountered (greater than 10), a sample of the new or missing DUs will be selected, and the sample weight will be adjusted accordingly. For more information, refer to Appendix C.

2. Extending the Coordinated 5-Year Sample

As was mentioned previously, the sample design was developed simultaneously for each of the 2005 through 2009 National Surveys on Drug Use and Health (NSDUHs), and the design was extended for the 2010-2011 and 2012-2013 NSDUHs. Starting with a census block-level frame, first- and second-stage sampling units (census tracts and area segments, respectively) were formed. A sufficient number of segments then were selected within sampled census tracts to support the 5-year design, several years beyond 2009, and any supplemental studies the Substance Abuse and Mental Health Services Administration (SAMHSA) chose to field.

2.1 Formation of and Objectives for Using the Composite Size Measures

The composite size measure procedure is used to obtain self-weighting[18] samples for multiple domains in multistage designs. The NSDUH sample design has employed the composite size measure methodology since 1988. The goal was to specify size measures for sample areas (segments) and dwelling units (DUs) that would achieve the following objectives:

- Yield the targeted domain sample sizes in expectation (E_s) over repeated samples; that is, if m_{ds} is the domain d sample size achieved by sample s, then

$$E_s(m_{ds}) = m_d \text{ for } d = 1,...,D. \quad (1)$$

- Constrain the maximum number of selections per DU at a specified value; specifically, the total number of within-DU selections was limited across all age groups to a maximum of 2.
- Minimize the number of sample DUs that must be screened to achieve the targeted domain sample sizes.
- Eliminate all variation in the sample inclusion probabilities within a domain, except for the variation in the within-DU/within-domain probabilities of selection. The inverse probabilities of selection for each sample segment were used to determine the number of sample DUs to select from within each segment. As a consequence, all DUs within a specific stratum were selected with approximately the same probability and, therefore, approximately equalized DU sampling weights. This feature minimizes the variance inflation that results from unnecessary variation in sampling weights.
- Equalize the expected number of sample persons per cluster to balance the interviewing workload and to facilitate the assignment of interviewers to regions and segments. This feature also minimizes adverse effects on precision resulting from extreme cluster size variations.
- Simplify the size measure data requirements so that census data (block-level counts) are adequate to implement the method.

[18] Self-weighting implies equal weights within domains defined by State and age group.

Using the 2000 census data supplemented with revised population projections, a composite size measure was computed for each census block defined within the United States. The composite size measure began by defining the rate $f_h(d)$ at which each age group domain d (d = 1,...,5 for 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 years or older) was to be sampled from State h.

Let $C_{hijk}(d)$ be the population count from domain d in census block k of segment j of State sampling (SS) region i within each State h. The composite size measure for block k was defined as

$$S_{hijk} = \sum_{d=1}^{5} f_h(d) \quad C_{hijk}(d). \tag{2}$$

The composite size measure for segment j was calculated as

$$S_{hij+} = \sum_{d=1}^{5} f_h(d) \sum_{k=1}^{N_{hij}} C_{hijk}(d), \tag{3}$$

where N_{hij} equals the number of blocks within segment j of SS region i and State h.

2.2 Stratification

Because the NSDUH design provides for estimates by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification. The objective of the next level of stratification was to distribute the number of interviews, in expectation, equally among SS regions. Within each State, census tracts were joined to form mutually exclusive and exhaustive SS regions of approximately equal sizes. Prior to forming the SS regions, composite size measures were scaled so that the aggregate composite size measure was roughly 100 per region. This scaling made it easier for the technician when forming the regions. Without scaling, the composite size measures would sum to approximately 75 (the expected sample size per region). Using desktop computer mapping software, the regions were formed, taking into account geographical boundaries, such as mountain ranges and rivers, to the extent possible. Therefore, the resulting regions facilitated ease of access and distributed the workload evenly among regions. A total of 900 SS regions were formed for the coordinated 2005-2009 design, and these strata definitions will remain the same through the 2013 survey.

In each of the 43 small States, which include the District of Columbia (see footnote 3), 12 SS regions were formed; however, 48 SS regions were formed in each of the 8 large States (i.e., in California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas). The number of SS regions to create in the small States versus the large States was determined as follows. The design called for 300 persons in each of three age groups (12 to 17, 18 to 25, and 26 or older) equally allocated to four quarters within each small sample State. Based on an analysis of the cost variance tradeoffs, an average cluster size of 3.125 persons in each of the three age groups (or an average of 9.375 persons over the three age groups combined) was considered near optimal. When applied to the small States, a quarterly sample of 75 persons per quarter per age group could be

obtained from 24 clusters or area segments. For unbiased variance estimation purposes, at least two observations are required per stratum (Chromy, 1981); maximum geographic stratification was obtained by defining 12 strata (SS regions) with 2 area segments each per quarter. Two additional segments were selected for each of the other three quarters, yielding 8 area segments per stratum, or 96 area segments per small sample State. This approach supported a target sample size for the small States of 300 persons per age group, or a total of 900 for the year. In the large sample States, 4 times as large a sample was required. Optimum cluster size configuration and maximum stratification given the need for unbiased variance estimation were maintained by simply quadrupling the number of strata (SS regions) to 48 per large sample State, yielding a sample 300 persons per age group per quarter, 1,200 per age group over four quarters, and 3,600 per year over all three age groups.

2.3 First- and Second-Stage Sample Selection

Once the SS regions were formed, the first-stage sampling units were created by collapsing adjacent census tracts within regions as needed. Although most census tracts contained 150 DUs in urban areas and 100 DUs in rural areas, some had to be collapsed in order to meet the minimum requirement. Once first-stage sampling units were formed, a probability proportional to the size sample was selected with minimum replacement within each SS region. The sampling frame was stratified implicitly by sorting the first-stage sampling units by a CBSA/SES (core-based statistical area/socioeconomic status) indicator and by the percentage of the population that is non-Hispanic and white. Table 2.1 summarizes the census tract sampling frame by State. In this table, a "census tract" is defined as one or more census tracts because some collapsing was done to meet the minimum size criteria.

To form segments within sampled census tracts, adjacent census blocks were collapsed until the total number of DUs within the area was at least 150 in urban areas and 100 in rural areas. In order to obtain geographic ordering of the blocks within tracts, block centroids were serpentine-sorted by latitude and longitude.[19] If a portion of a block fell between two other blocks but its centroid did not, the block was not combined with the other two blocks, and the resulting segment contained multiple pieces. However, the majority of segments consisted of contiguous blocks.

To control the geographic distribution of the sample, segments were sorted in the order they were formed, and one segment was selected per sampled census tract using the probability proportional to size sequential sampling method. As Table 2.1 indicates, 48 census tracts/segments per SS region were chosen for a total of 576 segments in each State, except in the large States where a total of 2,304 segments were chosen. Although only 24 segments per SS region were needed to support the 5-year study from 2005 through 2009, an additional 24 segments were selected to serve as replacements when segment DUs are depleted and/or to support any supplemental studies embedded within NSDUH. These 24 segments constitute the "reserve" sample and are available for use in 2010-2011 and 2012-2013 NSDUHs.

[19] The latitude and longitude for each census block were obtained from the Census 2000 Summary File 1.

Table 2.1 Number of Census Tracts and Segments on Sampling Frame, by State

State	State Abbreviation	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
Total U.S.			64,505	43,200	382,598		
Northeast							
Connecticut	CT	09	807	576	5,095	288	287
Maine	ME	23	343	576	3,533	288	287
Massachusetts	MA	25	1,355	576	6,163	288	288
New Hampshire	NH	33	272	576	3,076	288	286
New Jersey	NJ	34	1,914	576	5,657	288	288
New York	NY	36	4,738	2,304	19,057	1,152	1,149
Pennsylvania	PA	42	3,088	2,304	21,704	1,152	1,150
Rhode Island	RI	44	233	576	2,305	288	283
Vermont	VT	50	179	576	1,648	288	285
Midwest							
Illinois	IL	17	2,901	2,304	20,733	1,152	1,147
Indiana	IN	18	1,408	576	6,863	288	287
Iowa	IA	19	790	576	5,366	288	288
Kansas	KS	20	719	576	5,120	288	288
Michigan	MI	26	2,689	2,304	18,765	1,152	1,148
Minnesota	MN	27	1,293	576	5,955	288	288
Missouri	MO	29	1,303	576	7,193	288	287
Nebraska	NE	31	495	576	4,075	288	288
North Dakota	ND	38	215	576	1,618	288	279
Ohio	OH	39	2,902	2,304	20,342	1,152	1,149
South Dakota	SD	46	212	576	2,001	288	284
Wisconsin	WI	55	1,310	576	6,773	288	288
South							
Alabama	AL	01	1,079	576	6,958	288	288
Arkansas	AR	05	618	576	6,128	288	288
Delaware	DE	10	196	576	1,721	288	282
District of Columbia	DC	11	179	576	1,049	288	270
Florida	FL	12	3,140	2,304	25,374	1,152	1,150
Georgia	GA	13	1,609	576	7,682	288	288
Kentucky	KY	21	992	576	6,301	288	288
Louisiana	LA	22	1,099	576	5,841	288	288
Maryland	MD	24	1,204	576	5,477	288	288
Mississippi	MS	28	601	576	6,448	288	287
North Carolina	NC	37	1,550	576	8,708	288	287
Oklahoma	OK	40	977	576	5,654	288	286
South Carolina	SC	45	862	576	7,365	288	288
Tennessee	TN	47	1,246	576	7,534	288	288
Texas	TX	48	4,351	2,304	26,096	1,152	1,152
Virginia	VA	51	1,513	576	6,448	288	286
West Virginia	WV	54	466	576	4,319	288	287

(continued)

Table 2.1 Number of Census Tracts and Segments on Sampling Frame, by State (continued)

State	State Abbrevia-tion	State FIPS Code	Number of Census Tracts on Sampling Frame	Total Number of Census Tracts/Segments Selected	Number of Segments on Sampling Frame	Number Selected for 5-Year Sample	Unique Segments in 5-Year Sample
West							
Alaska	AK	02	154	576	1,348	288	283
Arizona	AZ	04	1,089	576	6,759	288	287
California	CA	06	6,978	2,304	22,973	1,152	1,152
Colorado	CO	08	1,050	576	6,231	288	288
Hawaii	HI	15	274	576	1,784	288	285
Idaho	ID	16	277	576	3,224	288	285
Montana	MT	30	256	576	2,417	288	288
Nevada	NV	32	474	576	3,919	288	288
New Mexico	NM	35	431	576	3,839	288	288
Oregon	OR	41	752	576	6,219	288	288
Utah	UT	49	485	576	4,024	288	288
Washington	WA	53	1,312	576	6,425	288	288
Wyoming	WY	56	125	576	1,291	288	283

FIPS = Federal information processing standards.

2.4 Survey Year and Quarter Assignment

The 48 sampled segments per SS region were randomly assigned to survey years by drawing equal probability subsamples of 4 segments. Prior to selecting the second subsample, the first subsample segments were removed from the pool of eligible segments. The second subsample then was selected from the remaining segments. This process was initially repeated 5 times until the 48 sampled segments were assigned to 6 subsamples of 4 and a "reserve" sample of 24 segments; for the 2010-2011 surveys, 2 additional subsamples of 4 segments were selected from the reserve sample. Similarly, for the 2012-2013 surveys, 2 more subsamples of 4 segments were chosen from the reserve sample.

The first subsample of segments was assigned to the 2005 NSDUH and constituted the panel of segments to be used for that year only. The second subsample of segments was assigned to the 2005 NSDUH and was used again in the 2006 survey; the third was assigned to the 2006 and 2007 surveys; and so on. Within each subsample, segments were assigned to survey quarters 1 through 4 in the order that they were selected.

Using the survey year and quarter assignments, a sequential segment identification number (SEGID) then was assigned. Table 2.2 describes the relationship between SEGIDs and quarter assignment. The last two digits in the SEGID are called the "segment suffix." The 2012 main survey corresponds to segment suffixes 29 through 36.

2.5 Creation of Variance Estimation Strata and Replicates

The nature of the stratified, clustered sampling design requires that the design structure be taken into consideration when computing variances of survey estimates. Key nesting variables

Table 2.2 Segment Identification Number Suffixes and Quarter Assignment

Segment Suffix	2005 NSDUH	2006 NSDUH	2007 NSDUH	2008 NSDUH	2009 NSDUH	2010 NSDUH	2011 NSDUH	2012 NSDUH	2013 NSDUH	Variance Replicate
01	x (Q1)									1
02	x (Q2)									1
03	x (Q3)									1
04	x (Q4)									1
05	x (Q1)	x (Q1)								2
06	x (Q2)	x (Q2)								2
07	x (Q3)	x (Q3)								2
08	x (Q4)	x (Q4)								2
09		x (Q1)	x (Q1)							1
10		x (Q2)	x (Q2)							1
11		x (Q3)	x (Q3)							1
12		x (Q4)	x (Q4)							1
13			x (Q1)	x (Q1)						2
14			x (Q2)	x (Q2)						2
15			x (Q3)	x (Q3)						2
16			x (Q4)	x (Q4)						2
17				x (Q1)	x (Q1)					1
18				x (Q2)	x (Q2)					1
19				x (Q3)	x (Q3)					1
20				x (Q4)	x (Q4)					1
21					x (Q1)	x (Q1)				2
22					x (Q2)	x (Q2)				2
23					x (Q3)	x (Q3)				2
24					x (Q4)	x (Q4)				2
25						x (Q1)	x (Q1)			1
26						x (Q2)	x (Q2)			1
27						x (Q3)	x (Q3)			1
28						x (Q4)	x (Q4)			1
29							x (Q1)	x (Q1)		2
30							x (Q2)	x (Q2)		2
31							x (Q3)	x (Q3)		2
32							x (Q4)	x (Q4)		2
33								x (Q1)	x (Q1)	1
34								x (Q2)	x (Q2)	1
35								x (Q3)	x (Q3)	1
36								x (Q4)	x (Q4)	1
37									x (Q1)	2
38									x (Q2)	2
39									x (Q3)	2
40									x (Q4)	2

Note: The segment suffix is defined as the last two digits of the segment identification number.

representing the variance estimation strata and replicates were created to capture explicit stratification and to identify clustering. For the 2005 through 2013 NSDUHs, variance estimation strata are defined at the SS region level, and each SS region appears in a different stratum every quarter. Because both census tracts and segments are nested within variance replicates, the variance contributions of both sampling units are covered by the nesting variables. Because one segment is selected per sampled census tract, the selection of census tracts at the first stage of selection may reduce variance by controlling the sample distribution and minimizing the chance of selecting neighboring and possibly similar segments within the same census tract.

To define the variance estimation strata for the 2005 through 2013 NSDUHs, the 900 SS regions were first placed in random order (States were randomly sorted, and regions were

randomly sorted within States). This list, numbered 1 to 900, defined the quarter 1 variance estimation strata (VESTRQ1). For quarter 2, the variance estimation strata, VESTRQ2, were defined as VESTRQ1 – 150 (or VESTRQ1 – 150 + 900 if VESTRQ1 is ≤ 150). Similarly, VESTRQ3 = VESTRQ2 – 150 (+ 900 if VESTRQ2 ≤ 150), and VESTRQ4 = VESTRQ3 – 150 (+ 900 if VESTRQ3 ≤ 150). As an example, an SS region that was assigned to stratum 151 in quarter 1 was assigned to stratum 1 (= 151 – 150) in quarter 2, stratum 751 (= 1 – 150 + 900) in quarter 3, and stratum 601 (= 751 – 150) in quarter 4. This method had the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SS regions from four different States.

The 2005 through 2013 definition of variance estimation strata has the effect of increasing the number of degrees of freedom for State-level estimates while preserving the number of degrees of freedom for national estimates (900). Each small sample State is in 48 different strata (12 SS regions × 4 quarters); therefore, there are 48 degrees of freedom available for State estimates. Similarly, each large sample State is in 192 strata (48 SS regions × 4 quarters) and therefore has 192 degrees of freedom for estimation.

Two replicates per year were defined within each variance stratum. Each variance replicate consists of four segments, one for each quarter of data collection. The first replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The second replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. Table 2.2 describes the assignment of segments to variance estimation replicates that are designed to account for positive covariance among consecutive year change estimates.

In addition to variance estimation strata and replicates, a sample weight is computed for each final respondent (see Section 3.9.1). The use of sample weights in analyses of NSDUH data is necessary to properly represent the target population and to account for disproportionate sampling by age group. All weighted statistical analyses for which variance estimates are needed should use the stratum and replicate variables to identify nesting. Variance estimates can be computed using a clustered data analysis software package such as SUDAAN® (RTI International, 2008). The SUDAAN software package computes variance estimates for nonlinear statistics using such procedures as a first-order Taylor series approximation of the deviations of estimates from their expected values. The approximation is unbiased for sufficiently large samples. SUDAAN also recognizes positive covariance among estimates involving data from 2 or more years.[20] Using data from the 2007 and 2008 NSDUHs and examining multiple measures, the average relative change in the standard error (SE) after accounting for covariance was about 1 percent.

2.6 Other Sampling-Related Variables

Because area segments consist of one or more census blocks, a number of demographic and geographic variables are available for sampled areas. The demographic data include the following: population counts by age, race, and ethnicity; estimated civilian, noninstitutional population aged 12 or older; DU counts; estimated group quarters units; and group quarters population by type of

[20] Using the variance estimation strata and replicates, SUDAAN recognizes positive covariance among estimates from consecutive years. For nonconsecutive years, strata are treated as collapsing with zero covariance.

group quarter.[21] For these variables, the block-level data were aggregated to form segment-level estimates.

The U.S. Census Bureau also makes available several geographic variables that can be associated with the 2005 through 2013 NSDUH sample segments. These are State, county and county name, census tract (equal to the most frequently occurring tract if the segment crosses multiple tracts), place name, census division and region, land area, CBSA/SES indicator (as defined in Section 2.3), county-level population density, and a rural or urban indicator.[22] Each census block is assigned a rural or urban status based on population density and/or proximity to a census-designated urbanized area (UA) or urban cluster (UC). In the NSDUH sample, if one or more of the blocks within a segment is urban, the segment is defined as urban. If 100 percent of the blocks are rural, the segment is defined as rural. Defining rural or urban status in this way provides an aggregate variable that is needed for assigning minimum size requirements (see Section 2.3) and for some data analyses. However, the definition slightly overestimates the urban population.

The 2005 through 2013 NSDUH sample was designed to facilitate matching to external data at the census tract level. Because field enumeration of the sample segments occurs at the segment level rather than the block level (see Section 3.3.1), only the group of blocks in which a NSDUH respondent resides is known. Therefore, specific census blocks are not linked to respondents or listing units. Because there is no direct linking of listing units to specific census blocks, no mechanism currently is available for assigning block-level data to NSDUH respondents. For this reason, some variables (e.g., place name and rural or urban indicator) may have some error associated with them.

[21] Data were obtained or derived from the Census 2000 Summary File 1 and adjusted using revised population counts from Claritas.

[22] All variables were obtained or derived from the Census 2000 Summary File 1.

3. General Sample Allocation Procedures for the Main Study

In this chapter, the computational details of the procedural steps used to determine both person and dwelling unit (DU) sample sizes are discussed. The within-DU age group-specific selection probabilities for the design of the 2012 National Survey on Drug Use and Health (NSDUH) also are addressed. This optimization procedure was designed specifically to address the Substance Abuse and Mental Health Services Administration's (SAMHSA's) multiple precision and design requirements while simultaneously minimizing the cost of data collection. Costs were minimized by determining the smallest number of interviews and selected DUs necessary to achieve the various design requirements. In summary, this three-step optimization procedure proceeded as follows:

1. In the first step, the optimal number of interviews (i.e., responding persons) by domains of interest needed to satisfy the precision requirements was determined for several drug use outcome measures. In other words, 255 unknown m_{ha} values for each State h (51) and age group a (5) were initially sought to be determined. A solution to this multiple constraint optimization was achieved using Chromy's Algorithm (Chromy, 1987). This is described in further detail in Section 3.2.
2. Using the m_{ha} determined from Step 1, the next step was to determine the optimal number of selected dwelling (D_{hj}) units (i.e., third-stage sample) that were necessary. This step was achieved by applying parameter constraints (e.g., probabilities of selection and expected response rates) at the segment level j or the stage at which DUs would be selected, which was done on a quarterly basis using approximately 25 percent of the m_{ha} values. This step is described in further detail in Section 3.3.
3. The final step in this procedure entailed determining age group-specific probabilities of selection (S_{hja}) for each segment given the m_{ha} and D_{hj} from Steps 1 and 2. This was achieved using a modification of Brewer's Method of Selection (Cochran, 1977, pp. 261-263). The modification was designed to select 0, 1, or 2 persons from each DU.[23] A detailed discussion of the final step is given in Section 3.4. After calculating the required DUs and the selection probabilities, sample size constraints were applied to ensure adequate samples for supplemental studies and to reduce the field interviewer (FI) burden. Limits on the total number of expected interviews per segment also were applied. This process became iterative to reallocate the reduction in sample size to other segments not affected by such constraints. Details of this step in the optimization procedure are given in Section 3.5.

[23] Direct application of Brewer's method would require a fixed sample size.

3.1 Notation

h = 50 States plus the District of Columbia.

a = Age group a = 1,...,5 and represents the following groups: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older.

j = Individual segment indicator (total of 7,200; 1,800 per quarter).

m_{ha} = Number of completed interviews (person respondents) desired in each State h and age group a. Computation of m_{ha} is discussed in Section 3.2. For quarterly computation of selected DU sample size, approximately 25 percent of the yearly estimate is used.

y_{ha} = Estimated number of persons in the target population in State h and age group a. The 2012 population is estimated using the 2000 census data adjusted to the 2007 Claritas population projections in the compound interest formula, $y = Ae^{Bx}$, where

y = population at time x,
A = initial population,
e = base of the system of natural logarithms,
B = growth rate per unit of time, and
x = period of time over which growth occurs.

First, B is computed as $[ln(y/A)]/x$, where y = the population in 2007, A = the population in 2000, and x = 7. Then the 2012 population (y^*_{ha}) is computed using the original formula and this time allowing x to be 12. Finally, the 2012 population is adjusted by the ratio of estimated eligible listed DUs to the Claritas DU counts (U_{hj}). This adjustment factor considers the number of added DUs expected to be obtained through the half-open interval rule (1.01) and the probability of a DU being eligible (ε_h), both determined via historic data. The coefficient adjustment of 1.01 is estimated using historical data and is the proportion of all screened DUs (includes added DUs) over the original total of selected DUs (excluding added DUs). So, $y_{ha} = \{[1.01 * \varepsilon_h * L_{hj} * (1/I_{hj}) / U_{hj}]\} * y^*_{ha}$, where ε_h , L_{hj} , and I_{hj} are defined further below. This adjustment is computed at the census block level and then aggregated to the State level.

f_{ha} = m_{ha} / y_{ha}. State-specific age group sampling fraction.

F_h = $Max[f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha}), a = 1\text{-}5]$.

P_{hj} = Inverse of the segment selection probability (includes the census tract selection probability). DU sample sizes are computed on a quarterly basis, and segments are selected on a yearly basis. Because each quarter contains only a fourth of the selected segments, these probabilities are adjusted by a factor of 4 so that weights will add to the yearly totals.

I_{hj} = Subsegmentation inflation factor. For segments too large to count and to list efficiently in both time and cost, field listing personnel may request that a portion of the segment be

randomly sampled. First, they perform a quick count (best guess: $L_1^*{}_{hj}$) of the entire segment. The sampling staff then subdivides the segment into roughly equal-sized subdivisions or subsegments (using a best guess estimate of the number of DUs in each subsegment: $B_1^*{}_{hj}$) and selects one for regular counting and listing. Beginning in 2008, some large segments were subsegmented based on census information prior to being sent to the field for listing. In some of these segments, the selected subsegment was still too large for listing, and a second round of subsegmenting was required. The second-level subsegmenting was performed in a similar fashion as the first-level subsegmenting, in that the first-level subsegment was counted (best guess: $L_2^*{}_{hj}$), and subdivided into roughly equal-sized subdivisions or subsegments (best guess: $B_2^*{}_{hj}$). Then, one subsegment was selected for regular counting and listing by sampling staff. For the subsegment to represent the entire segment, the weights are adjusted up to reflect the unused portion of the segment.

$= (B_1^*{}_{hj} / L_1^*{}_{hj})$, if one round of subsegmenting was done.

$= (B_1^*{}_{hj} / L_1^*{}_{hj}) * (B_2^*{}_{hj} / L_2^*{}_{hj})$, if two rounds of subsegmenting were required.

= 1, if no subsegmenting was done.

D_{hj} = Minimum number of DUs to select for screening in segment *j* to meet the targeted sample sizes for all age groups.

L_{hj} = Final segment count of DUs available for screening.

S_{hja} = State- and segment-specific probability of selecting a person in age group *a*. One implemented design constraint was that no single age group selection probability could exceed 1. The maximum allowable probability was then set to 0.99.

ε_h = State-specific DU eligibility rate. This rate was derived from 2010 NSDUH quarters 3 and 4 and 2011 NSDUH quarters 1 and 2 data by taking the average eligibility rate within each State.

ϕ_h = State-specific screening response rates. These rates were calculated using the same methodology as described for the DU eligibility rate (ε_h).

λ_{ha} = State- and age group-specific interview response rate. Using data from quarters 3 and 4 of the 2010 NSDUH and quarters 1 and 2 of the 2011 NSDUH, the additive effects of State and age group on interview response were determined by taking the average interview response rate within each State.

γ_{ha} = Expected number of persons within an age group per DU. This number was calculated using 2010 NSDUH quarters 3 and 4 and 2011 NSDUH quarters 1 and 2 data by dividing the weighted total number of rostered persons in an age group by the weighted total number of complete screened DUs by State.

δ_{ha} = State- and age group-specific maximum-of-two rule adjustment. The survey design restricts the number of interviews per DU to a total of two. This is achieved through a

modified Brewer's Method of Selection, which results in a loss of potential interviews in DUs where selection probabilities sum greater than 2. The adjustment is designed to inflate the number of required DUs to compensate for this loss. Using data from all four quarters of the 2010 NSDUH, the adjustment was computed by taking the average maximum-of-two rule adjustment within each State.

3.2 Determining Person Sample Sizes by State and Age Group

The first step in the design of the fourth stage of selection was to determine the optimal number of respondents needed in each of the 255 domains to minimize the costs associated with data collection, subject to multiple precision requirements established by SAMHSA. In summary, the precision requirements were that the expected relative standard error (RSE) on a prevalence of 10 percent not exceed the following:

- 3.00 percent for total population statistics, and
- 5.00 percent for statistics in three age group domains: 12 to 17, 18 to 25, and 26 or older.

In preparation for the 2005 through 2009 NSDUHs and the 2010-2011 and 2012-2013 NSDUHs, several optimization models and other related analyses were conducted. Using historical 2001 survey data, estimates and RSEs for each of nine outcome measures of interest were computed. Estimates then were standardized to a prevalence of 10 percent. The outcome measures of interest were included to address not only the NSDUH recency-of-use estimates, but also such related generic substance abuse measures as treatment received for alcohol and illicit drug use and dependence on alcohol and illicit drugs.

Specifically, the nine classes of NSDUH outcomes that were considered were as follows:

Use of Legal (Licit) Substances

1. *Cigarette Use in the Past Month.* Smoked cigarettes at least once within the past month.
2. *Alcohol Use in the Past Month.* Had at least one drink of an alcoholic beverage (beer, wine, liquor, or a mixed alcohol drink) within the past month.

Use of Illicit Substances

3. *Illicit Drug Use in the Past Month.* Includes use of hallucinogens, heroin, marijuana, cocaine, inhalants, opiates, or nonmedical use of sedatives, tranquilizers, stimulants, or pain relievers.
4. *Illicit Drug Use Other Than Marijuana in the Past Month.* Past month use of illicit drugs excluding those whose only illicit drug use was marijuana.
5. *Cocaine Use in the Past Month.* Use within the past month of cocaine in any form, including crack.

Note that current use of illicit drugs provides a broad measure of illicit drug use; however, it is dominated by marijuana and cocaine use. Therefore, estimates of illicit drug use other

than marijuana use and cocaine use are included because these two measures reflect different types of drug abuse.

Drug or Alcohol Dependence

6. *Dependent on Illicit Drugs in the Past Year.* Dependent on the same drugs listed in class 3, *Illicit Drug Use in the Past Month,* above. Those who are dependent on both alcohol and another illicit substance are included, but those who are dependent on alcohol only are not.
7. *Dependent on Alcohol and Not Illicit Drugs in the Past Year*. Dependent on alcohol and not dependent on illicit drugs.

Treatment for Drugs and Alcohol Problems

8. *Received Treatment for Illicit Drugs in the Past Year.* Received treatment in the past 12 months at any location (including hospitals, clinics, self-help groups, or doctors' offices) for illicit drug use.
9. *Received Treatment for Alcohol Use but Not Illicit Drug Use in the Past Year.* Received treatment in the past 12 months at any location (including hospitals, clinics, self-help groups, or doctors' offices) for drinking. These estimates exclude those who received treatment in the past 12 months for both drinking and illicit drug use.

These outcome measures, as well as the precision that is expected from this 2012 NSDUH design, are presented in Table 3.1, which was updated using 2010 NSDUH data. RSEs were based on an average prevalence rate of 10 percent for each measure.

Additionally, initial sample size requirements were implemented:

- Minimum sample size of 3,600 persons per State in the eight large States and 900 persons in the remaining 43 States.
- Equal allocation of the sample across the three age groups: 12 to 17, 18 to 25, and 26 or older within each State.

As in the 1999 through 2011 surveys, racial groups were not oversampled for the 2012 NSDUH. Consistent with previous surveys, the 2012 NSDUH was designed to oversample the younger age groups.

Among the 51 States, a required total sample size of 67,500 respondents was necessary to meet all precision and sample size requirements. Table 3.2 shows expected State by age group sample sizes. Because of the shorter calendar length of quarters 1 and 4 (due to interviewer training and the holidays, respectively), a decision was made to allocate the quarterly State by age group sample sizes (25 percent of the annual sample) to the four quarters in ratios of 96, 104, 104, and 96 percent, respectively. Only minor increases in unequal weighting resulted from not distributing the sample equally across quarters.

Table 3.1 Expected Relative Standard Errors, by Age Group

Outcome Measure	Total			12-17		
	Estimate	RSE	SRSE	Estimate	RSE	SRSE
Past Month Cigarette Use	22.97	1.34	2.20	8.34	3.09	2.79
Past Month Alcohol Use	51.79	0.75	2.34	13.59	2.43	2.89
Past Month Use of Illicit Drugs	8.92	2.16	2.02	10.13	2.85	2.87
Past Month Use of Illicit Drugs Other Than Marijuana	3.56	3.32	1.91	4.50	4.31	2.81
Past Month Cocaine Use	0.58	7.34	1.68	0.25	18.52	2.77
Past Year, Dependent on Illicit Drugs	1.90	4.23	1.77	2.53	5.46	2.64
Past Year, Dependent on Alcohol but Not Illicit Drugs	2.81	4.07	2.08	1.10	9.25	2.92
Past Year, Received Treatment for Illicit Drug Use	0.86	6.37	1.78	0.84	9.91	2.74
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.61	8.93	2.10	0.14	23.86	2.70
Average RSE	N/A	N/A	1.99	N/A	N/A	2.79
Target RSE	N/A	N/A	3.00	N/A	N/A	5.00
Outcome Measure	**18-25**			**26+**		
	Estimate	**RSE**	**SRSE**	**Estimate**	**RSE**	**SRSE**
Past Month Cigarette Use	34.24	1.37	2.97	22.83	1.68	2.74
Past Month Alcohol Use	61.53	0.81	3.07	54.85	0.88	2.92
Past Month Use of Illicit Drugs	21.51	1.85	2.90	6.57	3.44	2.74
Past Month Use of Illicit Drugs Other Than Marijuana	7.95	3.09	2.73	2.67	5.33	2.65
Past Month Cocaine Use	1.46	7.79	2.84	0.47	10.86	2.23
Past Year, Dependent on Illicit Drugs	5.37	4.13	2.95	1.22	7.70	2.56
Past Year, Dependent on Alcohol but Not Illicit Drugs	5.11	4.09	2.84	2.63	5.50	2.71
Past Year, Received Treatment for Illicit Drug Use	1.65	6.98	2.72	0.73	9.19	2.36
Past Year, Received Treatment for Alcohol Use but Not for Illicit Drug Use	0.87	10.02	2.82	0.62	11.10	2.64
Average RSE	N/A	N/A	2.87	N/A	N/A	2.62
Target RSE	N/A	N/A	5.00	N/A	N/A	5.00

RSE = relative standard error; SRSE = standardized relative standard error.

Table 3.2 Expected Main Study Sample Sizes, by State and Age Group

Region/State	State FIPS	SS Regions	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
Total Population		900	7,200	22,500	22,500	6,000	9,000	7,500	67,500
Northeast									
Connecticut	09	12	96	300	300	70	127	103	900
Maine	23	12	96	300	300	65	121	114	900
Massachusetts	25	12	96	300	300	77	124	99	900
New Hampshire	33	12	96	300	300	67	129	104	900
New Jersey	34	12	96	300	300	73	128	99	900
New York	36	48	384	1,200	1,200	312	491	397	3,600
Pennsylvania	42	48	384	1,200	1,200	283	479	438	3,600
Rhode Island	44	12	96	300	300	74	122	104	900
Vermont	50	12	96	300	300	68	120	112	900

(continued)

Table 3.2 Expected Main Study Sample Sizes, by State and Age Group (continued)

Region/State	State FIPS	SS Regions	Total Segments	Total Respondents					
				12-17	18-25	26-34	35-49	50+	Total
Midwest									
Illinois	17	48	384	1,200	1,200	332	489	379	3,600
Indiana	18	12	96	300	300	82	120	98	900
Iowa	19	12	96	300	300	76	117	107	900
Kansas	20	12	96	300	300	81	118	101	900
Michigan	26	48	384	1,200	1,200	297	487	416	3,600
Minnesota	27	12	96	300	300	81	122	97	900
Missouri	29	12	96	300	300	81	118	101	900
Nebraska	31	12	96	300	300	82	117	101	900
North Dakota	38	12	96	300	300	79	112	109	900
Ohio	39	48	384	1,200	1,200	307	478	415	3,600
South Dakota	46	12	96	300	300	79	113	108	900
Wisconsin	55	12	96	300	300	77	121	102	900
South									
Alabama	01	12	96	300	300	79	118	103	900
Arkansas	05	12	96	300	300	81	116	103	900
Delaware	10	12	96	300	300	75	121	104	900
District of Columbia	11	12	96	300	300	106	110	84	900
Florida	12	48	384	1,200	1,200	294	463	443	3,600
Georgia	13	12	96	300	300	85	128	87	900
Kentucky	21	12	96	300	300	81	119	100	900
Louisiana	22	12	96	300	300	84	117	99	900
Maryland	24	12	96	300	300	78	126	96	900
Mississippi	28	12	96	300	300	81	118	101	900
North Carolina	37	12	96	300	300	78	125	97	900
Oklahoma	40	12	96	300	300	84	114	102	900
South Carolina	45	12	96	300	300	76	120	104	900
Tennessee	47	12	96	300	300	80	120	100	900
Texas	48	48	384	1,200	1,200	359	496	345	3,600
Virginia	51	12	96	300	300	80	124	96	900
West Virginia	54	12	96	300	300	75	113	112	900
West									
Alaska	02	12	96	300	300	90	123	87	900
Arizona	04	12	96	300	300	87	117	96	900
California	06	48	384	1,200	1,200	343	500	357	3,600
Colorado	08	12	96	300	300	88	123	89	900
Hawaii	15	12	96	300	300	80	115	105	900
Idaho	16	12	96	300	300	87	115	98	900
Montana	30	12	96	300	300	76	111	113	900
Nevada	32	12	96	300	300	87	122	91	900
New Mexico	35	12	96	300	300	85	114	101	900
Oregon	41	12	96	300	300	83	114	103	900
Utah	49	12	96	300	300	108	113	79	900
Washington	53	12	96	300	300	83	120	97	900
Wyoming	56	12	96	300	300	84	112	104	900

FIPS = Federal information processing standards; SS = State sampling.

3.3 Third-Stage Sample Allocation for Each Segment

Given that the desired respondent sample size for each State and age group (m_{ha}) needed to meet the design parameters established by SAMHSA, the next step was to determine the minimal number of DUs to select for each segment to meet the targeted sample sizes. In short, this step involved determining the sample size of the third stage of selection. This sample size determination was performed on a quarterly basis to take advantage of both segment differences and, if necessary, make adjustments to design parameters. Procedures described below were developed originally for initial implementation in quarter 1 of the survey. The description is specific to quarter 1. Any modifications or corrections were made in subsequent quarters and are explained in detail in Section 3.8.

3.3.1 Dwelling Unit Frame Construction—Counting and Listing

The process by which the DU frame is constructed is called counting and listing. In summary, a certified lister visits the selected area and lists a detailed and accurate address (or description, if no address is available) for each DU within the segment boundaries. The lister is given a series of maps on which to mark the locations of these DUs. Map pages are formed so that the lister can easily navigate the segment and has sufficient space to denote the location of each DU. The number of map pages depends on the size and composition of the segment. In general, a sparsely populated rural segment has more map pages than a densely populated urban segment. Thus, segments in States like New York and Nevada have fewer map pages on average, while segments in States like South Dakota are much larger on average. The number of map pages per State and the average number of map pages per segment are summarized in Table 3.3. The list of DUs constructed during counting and listing is entered into a database and serves as the frame from which the third-stage sample is drawn.

In some situations, the number of DUs within the segment boundaries was much larger than the specified maximum. To obtain a reasonable number of DUs for the frame, the lister first counted the DUs in such an area. The sampling staff then partitioned the segment into smaller pieces or subsegments and randomly selected one to be listed. Beginning in 2008, some large segments were partitioned into subsegments using census information prior to being sent to the field. Sampling staff then randomly selected one subsegment to send to the field for listing. In a few of these cases, additional subsegmenting was required for one of the following reasons: (1) the area experienced high growth and the census counts used in the initial subsegment were outdated, or (2) there was not enough information available during the first subsegment, and the initial subsegment was still too large to list. Thus, an additional level of subsegmenting was implemented to make listing feasible. The number of segments that were subsegmented in the 2012 NSDUH sample is summarized in Table 3.4. For more information on the subsegmenting procedures, see Appendix B.

Table 3.3 Number of Map Pages, by State and Segment

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Total Population	7,200	41,465	5.8
Alabama	96	591	6.2
Alaska	96	583	6.1
Arizona	96	550	5.7
Arkansas	96	656	6.8
California	384	1,363	3.5
Colorado	96	500	5.2
Connecticut	96	385	4.0
Delaware	96	458	4.8
District of Columbia	96	285	3.0
Florida	384	1,861	4.8
Georgia	96	544	5.7
Hawaii	96	370	3.9
Idaho	96	775	8.1
Illinois	384	2,180	5.7
Indiana	96	562	5.9
Iowa	96	836	8.7
Kansas	96	696	7.3
Kentucky	96	548	5.7
Louisiana	96	605	6.3
Maine	96	629	6.6
Maryland	96	374	3.9
Massachusetts	96	426	4.4
Michigan	384	1,985	5.2
Minnesota	96	586	6.1
Mississippi	96	687	7.2
Missouri	96	631	6.6
Montana	96	988	10.3
Nebraska	96	892	9.3
Nevada	96	416	4.3
New Hampshire	96	542	5.6
New Jersey	96	413	4.3
New Mexico	96	967	10.1
New York	384	1,551	4.0
North Carolina	96	503	5.2
North Dakota	96	1,092	11.4
Ohio	384	1,905	5.0
Oklahoma	96	675	7.0
Oregon	96	513	5.3
Pennsylvania	384	2,394	6.2
Rhode Island	96	458	4.8
South Carolina	96	601	6.3
South Dakota	96	878	9.1
Tennessee	96	561	5.8
Texas	384	2,088	5.4

(continued)

Table 3.3 Number of Map Pages, by State and Segment (continued)

State	Total Segments	Cumulative Number of Map Pages per State	Average Number of Map Pages per Segment
Utah	96	502	5.2
Vermont	96	634	6.6
Virginia	96	405	4.2
Washington	96	576	6.0
West Virginia	96	620	6.5
Wisconsin	96	581	6.1
Wyoming	96	1,044	10.9

Table 3.4 Segment and Dwelling Unit Summary

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units	Sampled Dwelling Units	Added Dwelling Units
Total Population	7,200	644	2	1,543,451	212,746	1,528
Alabama	96	11	0	21,612	3,002	10
Alaska	96	13	0	21,565	2,388	36
Arizona	96	9	0	19,898	2,768	3
Arkansas	96	6	0	18,729	2,760	16
California	384	32	0	82,192	9,443	46
Colorado	96	8	0	20,921	3,059	12
Connecticut	96	8	0	21,042	2,832	23
Delaware	96	6	0	22,824	2,831	16
District of Columbia	96	20	1	29,800	4,974	81
Florida	384	66	0	87,798	12,734	34
Georgia	96	8	0	21,521	2,350	15
Hawaii	96	17	0	26,928	3,143	69
Idaho	96	6	0	18,235	2,297	3
Illinois	384	17	0	78,363	11,320	65
Indiana	96	4	0	18,432	2,482	9
Iowa	96	6	0	17,844	2,494	35
Kansas	96	6	0	18,700	2,577	21
Kentucky	96	5	0	21,453	2,828	24
Louisiana	96	7	0	21,122	2,731	10
Maine	96	8	0	20,528	3,807	59
Maryland	96	13	0	24,014	2,662	18
Massachusetts	96	11	0	20,890	3,014	50
Michigan	384	28	1	79,513	11,344	97
Minnesota	96	7	0	18,588	2,460	23
Mississippi	96	5	0	17,372	2,544	9
Missouri	96	10	0	19,043	2,855	24
Montana	96	15	0	17,084	3,281	14
Nebraska	96	6	0	17,758	2,545	11

(continued)

Table 3.4 Segment and Dwelling Unit Summary (continued)

State	Total Segments	Total Subsegmented Segments	Second-Level Subsegmented Segments	Listed Dwelling Units	Sampled Dwelling Units	Added Dwelling Units
Nevada	96	13	0	20,864	2,348	6
New Hampshire	96	4	0	19,775	2,905	85
New Jersey	96	7	0	21,390	2,608	14
New Mexico	96	13	0	18,605	2,768	3
New York	384	51	0	92,887	14,429	118
North Carolina	96	9	0	22,294	2,831	17
North Dakota	96	9	0	18,132	3,357	17
Ohio	384	23	0	80,173	11,664	58
Oklahoma	96	8	0	19,621	2,952	8
Oregon	96	7	0	19,902	2,535	12
Pennsylvania	384	23	0	78,670	11,843	64
Rhode Island	96	2	0	20,144	2,593	27
South Carolina	96	4	0	20,454	3,282	24
South Dakota	96	5	0	19,393	2,617	19
Tennessee	96	10	0	20,136	2,521	11
Texas	384	43	0	85,899	9,003	45
Utah	96	6	0	20,250	1,783	10
Vermont	96	5	0	19,440	3,250	42
Virginia	96	15	0	21,945	2,546	30
Washington	96	12	0	21,144	2,678	22
West Virginia	96	3	0	20,670	3,188	34
Wisconsin	96	5	0	17,308	2,431	9
Wyoming	96	9	0	20,586	3,089	20

During counting and listing, the lister moves about the segment in a prescribed fashion called the "continuous path of travel." Beginning from a starting point noted on the map,[24] the lister attempts to move in a clockwise fashion, makes each possible right turn, makes U-turns at segment boundaries, and does not break street sections. Within apartment buildings and group quarters, the lister attempts to apply the same rules; that is, the lister moves in a clockwise fashion and enumerates building floors from bottom to top. Following these defined rules and always looking for DUs on the right-hand side of the street (or hall), the lister minimizes the chance of not listing a DU within the segment. Also, using a defined path of travel makes it easier for the FI assigned to the segment to locate the sampled DUs. Finally, the continuous path of travel lays the groundwork for the half-open interval procedure for recovering missed DUs, as described in Section 3.7. A detailed description of the counting and listing procedures is provided in the 2012 counting and listing general manual (RTI International, 2011).

[24] Sampling staff review each map and determine the most logical starting point. They choose an intersection of two boundaries of the segment that seems most appropriate considering the segment's composition.

3.3.2 Determining Dwelling Unit Sample Size

For the main study, the optimization formula is as follows:

$$f_{ha} = P_{hj} * I_{hj} * (\frac{D_{hj}}{L_{hj}}) * S_{hja} * \phi_h * \lambda_{ha} * \delta_{ha}. \tag{4}$$

At this point in the procedure, only two components in the formula are unknown: D_{hj} and S_{hja}. Selection probabilities are segment- and age group-specific, and to maximize the number of selected persons within a DU, the age group whose adjusted sampling fraction [$f_{ha} / (\phi_h * \lambda_{ha} * \delta_{ha})$] = F_h, known now as the driving age group (see Section 1.5), is set to the largest allowable selection probability (S_{hja}) of 0.99. D_{hj} then is computed as

$$D_{hj} = \frac{f_{ha}}{(P_{hj} * I_{hj} * S_{hja} * \phi_h * \lambda_{ha} * \delta_{ha})} * L_{hj}. \tag{5}$$

3.4 Determining Fourth-Stage Sample (Person) Selection Probabilities for Each Segment

$$S_{hja} = \frac{f_{ha}}{P_{hj} * I_{hj} * (\frac{D_{hj}}{L_{hj}}) * \phi_h * \lambda_{ha} * \delta_{ha}}. \tag{6}$$

Having solved for D_{hj}, the selection probabilities for the remaining age groups were solved. If L_{hj} equals 0, D_h and S_{hja} are set to 0.

3.5 Sample Size Constraints: Guaranteeing Sufficient Sample for Additional Studies and Reducing Field Interviewer Burden

A major area of interest for the survey is to ensure that an adequate sample of eligible DUs remain within each segment. This sample surplus is needed to allow SAMHSA to implement supplemental studies if desired.

In addition, concern was noted about guaranteeing that FIs would be able to complete the amount of work assigned to them within the quarterly timeframe. These concerns prompted adjustments to the D_{hj} sample size:

1. Number of selected DUs for screening: < 100 or < ½L_{hj}. Adjustments were made by adjusting the D_{hj} counts to equal the minimum of 100 or ½L_{hj}.
2. Number of selected DUs: > 5. For cost purposes, if at least five DUs remain in the segment, the minimum number of selected DUs was set to five.
3. Expected number of interviews: < 40.

This expected number of interviews (m^*_{hja}) was computed as follows:

$$m^*_{hja} = D^*_{hj} * \varepsilon_h * \phi_h * \gamma_{ha} * S_{hja} * \lambda_{ha} * \delta_{ha}, \quad (7)$$

where D^*_{hj} has been adjusted for constraint 1. This value is the total number of interviews expected within each segment. The calculation of the first adjustment, the screening adjustment, is

$$5 / D^*_{hj} . \quad (8)$$

Similarly, the interview adjustment is computed as

$$40 / m^*_{hja} . \quad (9)$$

This second adjustment is applied to D_{hj} under the assumption of an equal number of screened DUs for each completed interview.

Both constraints 1 and 3 reduce the third-stage sample, which could in turn reduce the expected fourth-stage sample size. Therefore, the reduction in the third-stage sample is reallocated back to the segments by applying a marginal adjustment to the fourth-stage sample size (m_{ha}) at the State and age group level. As a result, segments that were not subject to these constraints could be affected. This adjustment to reallocate the DU sample is iterative until the expected person sample sizes are met.

3.6 Dwelling Unit Selection and Release Partitioning

After derivation of the required DU sample size within each State and segment (D_{hj}), the sample was selected from the frame of counted and listed DUs for each segment (L_{hj}). The frame was ordered in the same manner as described in Section 3.3.1, and selection was completed using systematic sampling with a random start value. Systematic sampling creates a heterogeneous sample of DUs by dispersing the sample throughout the segment. In addition, it minimizes social contagion from neighboring selected DUs that could have an impact on response rates and prevalence estimates. The listing order was used to approximate geographic location because a standard address is not available for all listed DUs.

To compensate for quarterly variations in response rates and yields, a sample partitioning procedure was implemented in all quarters. The entire sample (D_{hj}) still would be selected, but only certain percentages of the total would be released into the field. An initial percentage would be released in all segments at the beginning of the quarter. Based on interquarter work projections, additional percentages would be released 1 month into the quarter as needed and if field staff could handle the added workload. Each partitioning of the sample is a valid sample and helps manage the sample sizes by State without jeopardizing the validity of the study. Incidentally, a reserve sample of 20 percent also was selected, over and above the required D_{hj} sample, to allow for supplemental releases based on State experiences within each quarter. Thus, the 96 percent quarter 1 sample (see Section 3.2) was increased to the 115.2 percent level (i.e., 0.96 * 1.20 = 1.152). In quarter 1, the D_{hj} sample was allocated out to States in the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);

Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

As described in Section 3.9, a weight adjustment is applied to all DUs within a segment to account for the partial release of sample. The DU release adjustment (weight component #7) is equal to the inverse of the percentage of the sample that is released into the field. For example, if only DUs in release 1 were made available to the field, the DU release adjustment would equal 120/80 or 1.5. If releases 1, 4, and 5 were fielded, the adjustment would equal 120/100 or 1.2 because 80/120 + 10/120 + 10/120 = 100/120. A summary of the quarterly sample sizes and percentages released is provided in Table 3.5.

3.7 Half-Open Interval Rule and Procedure for Adding Dwelling Units

To guarantee that every DU had a chance of selection and to eliminate any bias associated with incomplete frames, a procedure called the half-open interval rule was implemented. This procedure required that the interviewer look both on the property of each selected DU and between that DU and the next listed DU for any unlisted units. When found in these specific locations, the unlisted units became part of the sample (added DUs). If the number of added DUs linked to any particular sample DU did not exceed 5, or if the number for the entire segment was less than or equal to 10, the FI was instructed to consider these DUs as part of his or her assignment. If either of these limits was exceeded, special subsampling procedures were implemented, as described in Appendix C. The number of added DUs in the 2012 NSDUH sample is summarized in the last column of Table 3.4.

Table 3.5 Quarterly Sample Sizes and Percentages Released

	Quarter 1			Quarter 2		
Region/State	**# Selected**	**# Released**	**Percentage**	**# Selected**	**# Released**	**Percentage**
Total Population	59,483	49,564	83	65,082	56,482	87
Northeast						
Connecticut	784	654	83	845	601	71
Maine	1,057	885	84	1,204	953	79
Massachusetts	972	805	83	800	762	95
New Hampshire	842	704	84	880	770	88
New Jersey	692	574	83	778	778	100
New York	4,100	3,413	83	4,731	3,942	83
Pennsylvania	2,952	2,456	83	3,886	3,556	92
Rhode Island	742	621	84	853	602	71
Vermont	972	811	83	991	863	87
Midwest						
Illinois	3,174	2,651	84	3,518	3,082	88
Indiana	717	597	83	759	659	87
Iowa	760	634	83	723	660	91
Kansas	739	616	83	832	624	75
Michigan	3,252	2,716	84	3,515	3,365	96
Minnesota	765	633	83	800	565	71
Missouri	726	610	84	804	803	100
Nebraska	698	580	83	783	655	84
North Dakota	902	752	83	1,019	934	92
Ohio	3,129	2,603	83	3,547	3,094	87
South Dakota	677	559	83	729	699	96
Wisconsin	770	642	83	825	651	79

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 1			Quarter 2		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
South						
Alabama	867	719	83	890	890	100
Arkansas	753	628	83	835	766	92
Delaware	838	695	83	908	683	75
District of Columbia	1,477	1,228	83	1,605	1,137	71
Florida	3,737	3,117	83	3,787	3,470	92
Georgia	641	535	83	692	606	88
Kentucky	707	592	84	835	835	100
Louisiana	732	614	84	823	749	91
Maryland	777	644	83	742	619	83
Mississippi	709	597	84	810	612	76
North Carolina	780	646	83	838	803	96
Oklahoma	685	567	83	960	801	83
South Carolina	815	680	83	901	862	96
Tennessee	696	583	84	827	721	87
Texas	2,545	2,124	83	2,675	2,338	87
Virginia	781	655	84	824	687	83
West Virginia	878	731	83	966	965	100
West						
Alaska	753	623	83	667	532	80
Arizona	895	745	83	857	573	67
California	2,690	2,237	83	2,971	2,347	79
Colorado	806	671	83	880	762	87
Hawaii	779	650	83	902	900	100
Idaho	629	527	84	603	603	100
Montana	902	755	84	967	844	87
Nevada	589	487	83	596	472	79
New Mexico	710	593	84	809	706	87
Oregon	769	644	84	726	667	92
Utah	476	394	83	508	485	95
Washington	834	695	83	939	628	67
Wyoming	811	672	83	917	801	87

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
Total Population	64,283	55,816	87	59,644	50,884	85
Northeast						
Connecticut	824	823	100	754	754	100
Maine	1,154	963	83	1,006	1,006	100
Massachusetts	770	769	100	712	678	95
New Hampshire	889	782	88	860	649	75
New Jersey	898	671	75	741	585	79
New York	4,379	3,650	83	4,103	3,424	83
Pennsylvania	3,758	3,124	83	3,423	2,707	79
Rhode Island	866	758	88	868	612	71
Vermont	1,093	911	83	889	665	75
Midwest						
Illinois	3,562	3,110	87	3,134	2,477	79
Indiana	755	628	83	716	598	84
Iowa	775	651	84	692	549	79
Kansas	767	731	95	694	606	87
Michigan	3,463	2,892	84	3,167	2,371	75
Minnesota	706	647	92	615	615	100
Missouri	873	685	78	790	757	96
Nebraska	844	703	83	811	607	75
North Dakota	953	913	96	911	758	83
Ohio	3,432	3,004	88	3,232	2,963	92
South Dakota	728	637	88	722	722	100
Wisconsin	754	596	79	689	542	79

(continued)

Table 3.5 Quarterly Sample Sizes and Percentages Released (continued)

Region/State	Quarter 3			Quarter 4		
	# Selected	# Released	Percentage	# Selected	# Released	Percentage
South						
Alabama	807	771	96	783	622	79
Arkansas	839	668	80	765	698	91
Delaware	947	829	88	792	624	79
District of Columbia	1,544	1,285	83	1,670	1,324	79
Florida	3,692	3,082	83	3,501	3,065	88
Georgia	670	640	96	717	569	79
Kentucky	814	711	87	821	690	84
Louisiana	809	702	87	799	666	83
Maryland	742	708	95	691	691	100
Mississippi	831	758	91	817	577	71
North Carolina	815	712	87	806	670	83
Oklahoma	905	828	91	823	756	92
South Carolina	942	942	100	833	798	96
Tennessee	775	584	75	760	633	83
Texas	2,628	2,197	84	2,452	2,344	96
Virginia	737	581	79	652	623	96
West Virginia	992	663	67	829	829	100
West						
Alaska	672	644	96	589	589	100
Arizona	774	646	83	804	804	100
California	3,060	2,681	88	2,611	2,178	83
Colorado	911	911	100	865	715	83
Hawaii	993	869	88	913	724	79
Idaho	642	642	100	604	525	87
Montana	882	882	100	800	800	100
Nevada	838	734	88	824	655	79
New Mexico	830	796	96	734	673	92
Oregon	709	652	92	684	572	84
Utah	461	459	100	483	445	92
Washington	863	749	87	809	606	75
Wyoming	916	842	92	884	774	88

3.8 Quarter-by-Quarter Deviations

This section describes corrections and/or modifications that were implemented in the process of design optimization. "Design" refers to deviations from the original proposed plan of design. "Procedural" refers to changes made in the calculation methodologies. Finally, "Dwelling Unit Selection" addresses changes that occurred after sample size derivations, specifically corrections implemented during fielding of the sample (i.e., sample partitioning as described in Section 3.6). Quarter 1 deviations are not included because the methods and procedures described above were all implemented in quarter 1. Subsequently, any changes would have been made after quarter 1.

Quarter 2

Design: An additional 20 percent reserve sample was added to the 104 percent quarterly sample to allow for supplemental releases where needed. Thus, the total quarter 2 sample was increased to the 124.8 percent level.

Procedural: To predict State response rates more accurately, the most current four quarters of data were used in the computation of State-specific yield and response rates. Thus, data from quarters 1 through 4 of the 2011 NSDUH were used to compute average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 2 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 3

Design: Using the completed cases from quarter 1 and the projected number of completes from quarter 2, each State's midyear surplus/shortfall was computed. The quarter 3 104 percent sample then was adjusted by this amount. An additional 20 percent sample also was included, bringing the total quarter 3 adjusted sample to the 124.8 percent level.

Procedural: Data from quarters 2 through 4 of the 2011 NSDUH and quarter 1 of the 2012 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 3 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

Quarter 4

Design: The State and age 96 percent quarterly sample sizes were adjusted to meet the yearly targets based on completed cases from quarters 1 and 2 and the projected number of completes from quarter 3. An additional 20 percent sample also was included, bringing the total quarter 4 adjusted sample to the 115.2 percent level.

Procedural: Data from quarters 3 and 4 of the 2011 NSDUH and quarters 1 and 2 of the 2012 NSDUH were used to compute State-specific average yields, DU eligibility, screening response, and interviewer response rates.

Dwelling Unit Selection: The quarter 4 D_{hj} sample was partitioned into the following release percentages:

Release 1: 67 percent of entire sample (80/120, main sample + 20 percent reserve);
Release 2: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 3: 4 percent of entire sample (5/120, main sample + 20 percent reserve);
Release 4: 8 percent of entire sample (10/120, main sample + 20 percent reserve);
Release 5: 8 percent of entire sample (10/120, main sample + 20 percent reserve); and
Release 6: 8 percent of entire sample (10/120, main sample + 20 percent reserve).

3.9 Sample Weighting Procedures

At the conclusion of data collection for the last quarter, design weights are constructed for each quarter of the State-level study, reflecting the various stages of sampling. At the time this report was published, the weights for the 2012 NSDUH had not yet been computed. However, the planned procedures are described in this section.

3.9.1 Main Study Sampling Weights

The calculation of the sampling weights will be based on the stratified, four-stage design of the study. Specifically, the person-level sampling weights will be the product of the four stagewise sampling weights, each equal to the inverse of the selection probability for that stage. In review, the stages are as follows:

Stage 1: Selection of census tract.

Stage 2: Selection of segment.

Stage 3: Selection of DU.

Three possible adjustments exist with this stage of selection:

(1) subsegmentation inflation: by-product of counting and listing (includes up to two levels of subsegmenting);
(2) added DU: results from the half-open interval rule when subsampling is needed; and
(3) release adjustment.

Stage 4: Selection of person within a DU.

A total of seven weight adjustments will be necessary for the calculation of the final analysis sample weight. All weight adjustments will be implemented using a generalized exponential model (GEM) technique. These adjustments are listed in the order in which they will be implemented:

1. *Nonresponse Adjustment at the Dwelling Unit Level.* This adjustment is to account for the failure to complete the within-DU roster. The potential list of variables for the 51-State main study DU nonresponse modeling is presented in Table 3.6.
2. *Dwelling Unit-Level Poststratification.* This adjustment involves using screener data of demographic information (e.g., age, race, gender). DU weights will be adjusted to the intercensal population estimates derived from the 2000 census for various demographic domains. In short, explanatory variables used during modeling will consist of counts of eligible persons within each DU that fall into the various demographic categories. Consequently, these counts, multiplied by the newly adjusted DU weight and summed across all DUs for various domains, will add to the census population estimates. This adjustment is useful for providing more stable control totals for subsequent adjustments and pair weights. Potential explanatory variables are listed in Table 3.7.
3. *Extreme Weight Treatment at the Dwelling Unit Level.* If it is determined that design-based weights (stages 1 and 2) along with any of their respective adjustments result in an unsatisfactory unequal weighting effect (i.e., variance of the DU-level weights is too high, with high frequency of extreme weights), then extreme weights will be further adjusted. This adjustment will be implemented by doing another weight calibration. The control totals are the DU-level poststratified weights, and the same explanatory variables as in DU-level poststratification will be used so that the extreme weights are controlled and all the distributions in various demographic groups are preserved.
4. *Selected Person Weight Adjustment for Poststratification to Roster Data.* This step utilizes control totals derived from the DU roster that are already poststratified to the census population estimates. This adjustment assists in bias reduction and improved precision by taking advantage of the properties of a two-phase design. Selected person sample weights (i.e., those that have been adjusted at the DU level and account for fourth-stage sampling) are adjusted to the DU weight sums of all eligible rostered persons. Any demographic information used in modeling is based solely on screener information because this is the only information available for all rostered persons. Potential explanatory variables for this adjustment are a combination of the variables presented in Table 3.8.
5. *Person-Level Nonresponse Adjustment.* This adjustment allows for the correction of weights resulting from the failure of selected sample persons to complete the interview. Respondent sample weights will be adjusted to the weight of all selected persons. Again, demographic information used in modeling is based solely on screener information. Potential explanatory variables for this adjustment are a combination of the variables presented in Table 3.8.
6. *Person-Level Poststratification.* This step is to adjust the final person sample weights to the census population estimates derived from the 2000 census. These are the same outside control totals used in the second adjustment. However, demographic variables for this adjustment are based on questionnaire data, not screener data as in adjustments 2, 4, and 5. Potential explanatory variables used in modeling are presented in Table 3.7.
7. *Extreme Weight Treatment at the Person Level.* This adjustment will be implemented in the same manner as described in adjustment 3, except that the weights reflect the fourth stage of selection.

All weight adjustments for the 2012 main study's final analysis weights will be derived from a GEM technique. To help reduce computational burden at all adjustment steps, separate models will be fit for clusters of States, based on census division definitions as shown in Table 3.9. Furthermore, model variable selection at each adjustment will be done using a combination method of forward and backward selection processes. The forward selection will be used for the model enlargement. Within each enlargement, backward selection will be used. The final adjusted weight, which is the product of weight components 1 through 15, is the analysis weight used in estimation. Exhibit 1 presents a flowchart of steps used in the weighting process, and Exhibit 2 displays all individual weight components.

Full details of the finalized modeling procedures, as well as final variables used in each adjustment step, will be described in the forthcoming report on the person-level sampling weight calibration for the 2012 NSDUH (Chen et al., in press).

Table 3.6 Definitions of Levels for Potential Variables for Dwelling Unit Nonresponse Adjustment

Group Quarters Indicator
1: College Dorm
2: Other Group Quarters
3: Nongroup Quarters
Percentage of Owner-Occupied Dwelling Units in Segment (% Owner)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Percentage of Blacks in Segment (% Black)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Percentage of Hispanics in Segment (% Hispanic)
1: 0 - <10%
2: 10% - <50%
3: 50% - 100%
Population Density
1: CBSA > 1,000,000
2: CBSA < 1,000,000
3: Non-CBSA Urban
4: Non-CBSA Rural
Quarter
1: Quarter 1
2: Quarter 2
3: Quarter 3
4: Quarter 4
Segment Combined Median Rent and Housing Value (Rent/Housing)
1: First Quintile
2: Second Quintile
3: Third Quintile
4: Fourth Quintile
5: Fifth Quintile
State

CBSA = core-based statistical area.

Note: Interactions among the main effect variables also are considered.

Table 3.7 Definitions of Levels for Potential Variables for Dwelling Unit Poststratification and Respondent Poststratification at the Person Level

Age
1: 12-17
2: 18-25
3: 26-34
4: 35-49
5: 50+[a]
Gender
1: Male
2: Female
Hispanicity
1: Hispanic
2: Non-Hispanic
Quarter
1: Quarter 1
2: Quarter 2
3: Quarter 3
4: Quarter 4
Race
1: White
2: Black
3: American Indian/Alaska Native
4: Asian
5: Two or More Races
State

Note: Interactions among the main effect variables also are considered.

[a] For person-level respondent poststratification adjustment, the age category of 50+ is further divided into the 50-64 and 65+ categories.

Table 3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment

Group Quarters Indicator
- 1: College Dorm
- 2: Other Group Quarters
- 3: Nongroup Quarters

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percentage of Blacks in Segment (% Black)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Percentage of Hispanics in Segment (% Hispanic)
- 1: 0 - <10%
- 2: 10% - <50%
- 3: 50% - 100%

Population Density
- 1: CBSA > 1,000,000
- 2: CBSA < 1,000,000
- 3: Non-CBSA Urban
- 4: Non-CBSA Rural

Quarter
- 1: Quarter 1
- 2: Quarter 2
- 3: Quarter 3
- 4: Quarter 4

Segment Combined Median Rent and Housing Value (Rent/Housing)
- 1: First Quintile
- 2: Second Quintile
- 3: Third Quintile
- 4: Fourth Quintile
- 5: Fifth Quintile

State

Age
- 1: 12-17
- 2: 18-25
- 3: 26-34
- 4: 35-49
- 5: 50+

Gender
- 1: Male
- 2: Female

(continued)

Table 3.8 Definitions of Levels for Potential Variables for Selected Person Poststratification and Person-Level Nonresponse Adjustment (continued)

Hispanicity
1: Hispanic
2: Non-Hispanic
Race
1: White
2: Black
3: American Indian/Alaska Native
4: Asian
5: Two or More Races
Relation to Householder
1: Householder or Spouse
2: Child
3: Other Relative
4: Nonrelative

CBSA = core-based statistical area.

Note: Interactions among the main effect variables also are considered.

Table 3.9 Model Group Definitions (Census Divisions)

Model	Defined State
1	Connecticut, Maine, New Hampshire, Rhode Island, Vermont, Massachusetts
2	New Jersey, New York, Pennsylvania
3	Illinois, Indiana, Michigan, Wisconsin, Ohio
4	Iowa, Kansas, Minnesota, Missouri, Nebraska, South Dakota, North Dakota
5	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
6	Alabama, Kentucky, Mississippi, Tennessee
7	Arkansas, Louisiana, Oklahoma, Texas
8	Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Arizona
9	Alaska, Hawaii, Oregon, Washington, California

Exhibit 1 Flowchart of Sample Weighting Steps

Exhibit 2 Sample Weight Components

Dwelling Unit-Level Design Weight Components	
#1.	Inverse Probability of Selecting Census Tract
#2.	Inverse Probability of Selecting Segment
#3.	Quarter Segment Weight Adjustment
#4.	Subsegmentation Inflation Adjustment
#5.	Inverse Probability of Selecting Dwelling Unit
#6.	Inverse Probability of Added/Subsampled Dwelling Unit
#7.	Dwelling Unit Release Adjustment
#8.	Dwelling Unit Nonresponse Adjustment
#9.	Dwelling Unit Poststratification Adjustment
#10.	Dwelling Unit Extreme Weight Adjustment
Person-Level Design Weight Components	
#11.	Inverse Probability of Selecting a Person within a Dwelling Unit
#12.	Selected Person Poststratification to Roster Adjustment
#13.	Person-Level Nonresponse Adjustment
#14.	Person-Level Poststratification Adjustment
#15.	Person-Level Extreme Weight Adjustment

3.9.2 Quality Control Measures in Design-Based Weighting Procedures

Quality control (QC) measures are applied to every component of the DU-level and person-level design weights. In addition to the QC measures outlined below, SAS programs are examined for errors, warnings, and uninitialization in the log by a sampling team member and reviewed by a different sampling team member. The following QC measures are employed to ensure the accuracy of design-based weight calculations:

- For segments that are subsegmented, check that the subsegmenting adjustment factor is greater than 1 (i.e., the count for the entire segment is greater than the count for the subsegment). This check is also performed for segments that are subsegmented twice.
- Compare the DU eligibility indicator with the completed screener indicator. Make sure all screener-complete DUs are eligible.
- Compare the final screening code with the DU eligibility and completed screener indicators to ensure these variables are defined correctly.
- Check the subsampling rate for added DUs that are subsampled. Review the frequency distribution of the DU subsampling rates to check values and ensure that the correct number of DUs are adjusted.
- Check that the minimum and maximum values of the DU release weight factor are within the expected range and that there are no missing values.
- Check the household-level weight to ensure that there are no missing values and the sum is close to the expected value.
- Compare the person-level indicators for eligible, selected, and complete. Make sure that all completed cases are selected and that all selected cases are eligible.
- Compare the final interview code with the person-level eligibility indicator to make sure this variable is defined correctly.
- Make sure that the probability of selection is nonmissing for all selected persons.
- Check the maximum-of-two selected persons' adjustment to make sure that the maximum value is two.
- Check the person-level weight to ensure that there are no missing values and the sum is close to the expected value.

4. General Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

As was done in the 2008 through 2011 National Surveys on Drug Use and Health (NSDUHs), a Mental Health Surveillance Study (MHSS) was embedded in the 2012 NSDUH. In 2008, a split-sample MHSS was conducted to calibrate the Kessler-6 (K6) nonspecific psychological distress scale and two competing functional impairment scales with a "gold-standard" clinical psychiatric interview in order to generate prevalence estimates of serious mental illness (SMI) among adults aged 18 or older. Based on the results from the 2008 MHSS, an abbreviated version of the World Health Organization-Disability Assessment Scale (WHODAS) (Novak, Colpe, Barker, & Gfroerer, 2010; Rehm et al., 1999) was adopted for the 2009 through 2012 surveys. The overarching goal of the MHSS was to fit a prediction model for SMI among adults aged 18 or older that can be used to create accurate model-based estimates of SMI prevalence at the national and domain levels (e.g., by age group and race/ethnicity). In addition, the MHSS data have the potential for a variety of important analyses beyond this primary purpose, including the model-free estimation of SMI prevalence and model-based and model-free estimation of moderate mental illness (MMI), mild mental illness (LMI), any mental illness (AMI), and specific mental disorders. The procedures for conducting the study remained the same. That is, a sample of respondents was contacted by phone for clinical follow-up.

4.1 Respondent Universe and Sampling Methods for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

The 2012 MHSS was designed to yield 500 clinical follow-up interviews during 2012. In addition to the MHSS, a random sample of 1,000 clinical interviews was conducted to further refine and adjust the calibration of the NSDUH questions to measure SMI. This study is called the Expanded Mental Health Surveillance Study (EMHSS). The total sample size for the 2012 MHSS and EMHSS is 1,500 completed interviews. The sample was embedded in the main study sample; therefore, the initial interview for the validation cases was included in the target of 45,000 main study adult interviews. The target population for the MHSS excluded persons whose main study interview was conducted in Spanish.

A new selection algorithm was developed for the 2010 MHSS to mitigate the problem of extreme weights seen in the 2008 MHSS and was continued in 2011 and 2012. A subsample of eligible respondents aged 18 or older was selected for clinical follow-up with probabilities based on their K6 nonspecific psychological distress scale score (Kessler et al., 2003) and WHODAS score, and an age group adjustment factor was applied. A probability sampling algorithm was programmed in the computer-assisted interviewing (CAI) instrument so that field interviewers (FIs) could recruit selected respondents for the subsequent clinical psychiatric interview conducted by telephone.

4.2 Person Sample Allocation Procedures for the Mental Health Surveillance Study and Expanded Mental Health Surveillance Study

Table 4.1 shows some of the age-related factors used to compute sampling rates. Based on 2010 population estimates and the 2012 planned sample, the average weighting for persons aged 50 or older is 8.5 times as large as the average weighting for persons aged 18 to 25. (Smaller differences occur for intermediate age groups.) To compensate for this initial disparity in weights and to focus on persons aged 18 or older as a whole, sampling rates were set for persons aged 18 to 25, then adjusted for the other three age groups by applying the equalization factor, *F*, shown in Table 4.1.[25] The eligibility of questionnaire respondents for the clinical follow-up was based on the language used to complete the questionnaire. Persons completing the Spanish-language questionnaire were not eligible to be selected for the clinical follow-up. Response rates shown are the product of the percentage agreeing to the follow-up survey and the proportion of those who actually participate.[26]

Table 4.1 MHSS/EMHSS Age-Related Factors

Age	2010 Population	Planned Sample	Average Weight	Weight Equalization Factor	Eligibility Factor	Response Rate Factor	Overall Age-Related Factor[1]
18 to 25	34,072,349	22,500	1,514	1.0000	97.08%	67.72%	1.0000
26 to 34	36,523,574	6,000	6,087	4.0198	93.03%	71.89%	3.9517
35 to 49	62,042,733	9,000	6,894	4.0198	94.64%	67.97%	4.1088
50 or Older	96,633,922	7,500	12,885	4.0198	96.76%	60.00%	4.5525

[1] The overall age-related factor is the weight equalization factor divided by the eligibility and response rate factors and then normalized.

Predictive models were developed to estimate the probability of SMI based on the K6 and WHODAS scores. In addition, the sample distribution by K6 and WHODAS scores was computed from the 2010 NSDUH data by the four sample allocation age groups represented in the 18 or older population.

The general sample allocation strategy was to find an allocation that provided the most precise estimate of SMI so that appropriate cut points could be established that produced the correct overall prevalence measure. A total of 225 strata were defined based on the combination of 25 possible K6 scores (0 to 24)[27] and 9 possible WHODAS scores (0 to 8). Predicted probabilities of SMI were used to obtain proportionality factors, $r_{h,age}$, for setting sampling rates by stratum (denoted *h*) and age group:

[25] Use of the derived weight equalization factors in Table 4.1 would have greatly increased the sampling rate for persons aged 50 or older. An adjusted set of factors that partially reduced the unequal weighting effects across age groups was specified instead. The adjusted equalization factors for the 35 to 49 and 50 or older age groups were set equal to the factor for the 26 to 34 age group.

[26] Eligibility and response rate factors were computed using 2010 MHSS data.

[27] In the prediction model, a recoded form of K6 score was used: scores 0 to 7 were recoded as 0, and all other scores had 7 subtracted from them to give a recoded total ranging from 0 to 17. These scores were reverse recoded to get back to the original K6 scores that were used in the two-way matrix. Thus, the predicted probabilities of mental illness are all identical for K6 scores of 0 to 7.

$$r_{h,age} \propto \frac{\sqrt{P_h(1-P_h)}}{E_{age}RR_{age}} * F_{age},$$

where P_h refers to the predicted probability of SMI in stratum h, and F_{age}, E_{age}, and RR_{age} refer to the age-specific weight equalization factors, eligibility factors, and response rate factors, respectively. These proportionality factors then were multiplied by the projected sample counts and scaled to achieve an overall respondent sample of 1,500 persons aged 18 or older to obtain the stratum and age-specific sampling rates. As an example, the predicted probability of SMI for a person with a K6 score of 10 and a WHODAS score of 6 was 0.1398. For the 18 to 25 age group, the proportionality factor then would be

$$r_{h,18-25} = \frac{\sqrt{0.1398(1-0.1398)}}{0.9708 * 0.6772} * 1.000 = 0.5275$$

An adjustment factor of 0.0916 was applied to each proportionality factor in order to achieve an overall sample of 1,500 persons. Thus, the sampling rate for this stratum and age group was 0.5275 * 0.0916 = 0.0483.

Projected yields of positive cases based on the predicted probability of SMI broken out by age group are provided in Table 4.2. In addition, Tables 4.3 and 4.4 provide the 2012 MHSS/EMHSS sample allocation by K6 group and WHODAS score, respectively.

Table 4.2 Projected Yields of Predicted Positive Cases

	Age Group				
	18 to 25	**26 to 34**	**35 to 49**	**50 or Older**	**18 or Older**
SMI	76	69	92	45	281
AMI	195	174	224	126	720
Total Sample	338	336	476	351	1,500

AMI = any mental illness, SMI = serious mental illness.

The probability sample of 1,500 clinical follow-up interviews was distributed across four calendar quarters with approximately 375 follow-up interviews per quarter. Throughout the 2012 survey, the MHSS/EMHSS sample was monitored and the sampling parameters were modified on an as-needed basis.

The 2012 MHSS/EMHSS resulted in 1,622 completed clinical interviews. An estimated 84 percent of selected persons agreed to participate in the MHSS/EMHSS, and 79 percent of those persons completed the clinical interview.

The analysis weight for the MHSS/EMHSS sample will be separate from the main study analysis weight. To compute the MHSS/EMHSS analysis weight, the main study analysis weight will be multiplied by the inverse of the probability of selecting the person for clinical follow-up (the inverse of the sampling parameter used to select the person) and nonresponse and poststratification adjustments (Exhibit 3).

Table 4.3 2012 MHSS/EMHSS Sample Allocation, by K6 Group

K6 Group	Percent of Population[1]	Assumed SMI Rate (Percent)[2,3]	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
0 to 3	53.54	0.92	497	0	0.02370
4 to 5	13.67	1.21	145	0	0.02386
6 to 7	9.08	1.58	114	0	0.02534
8 to 9	6.26	2.86	98	0	0.03126
10 to 11	4.41	5.51	99	4	0.04192
12 to 15	6.71	12.28	225	52	0.05731
16 or Higher	6.33	41.85	322	225	0.07980
Total	**100.00**	**4.70**	**1,500**	**281**	

K6 = Kessler-6, a 6-item psychological distress scale; SMI = serious mental illness.

[1] Source: 2010 National Survey on Drug Use and Health.

[2] Source: 2009 Mental Health Surveillance Study.

[3] To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across K6 scores. These rates are not the actual SMI rates that were used in the sample allocation.

Table 4.4 2012 MHSS/EMHSS Sample Allocation, by WHODAS Score

WHODAS Score	Percent of Population[1]	Assumed SMI Rate (Percent)[2,3]	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
0	74.26	1.06	738	0	0.02210
1	7.41	2.61	116	3	0.03473
2	4.95	4.67	104	4	0.04695
3	3.37	7.68	93	8	0.06132
4	2.61	13.45	89	22	0.07585
5	2.05	21.26	89	36	0.09634
6	1.83	32.44	87	57	0.10606
7	1.44	42.75	75	56	0.11673
8	2.10	58.81	108	95	0.11413
Total	**100.00**	**4.70**	**1,500**	**281**	

K6 = Kessler-6, a 6-item psychological distress scale; SMI = serious mental illness; WHODAS = World Health Organization-Disability Assessment Scale.

[1] Source: 2010 National Survey on Drug Use and Health.

[2] Source: 2009 Mental Health Surveillance Study.

[3] To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across K6 scores. These rates are not the actual SMI rates that were used in the sample allocation.

In addition to the analysis weight, MHSS/EMHSS-specific variance estimation strata and replicates will be created to appropriately account for the design when computing variances of estimates. To create the MHSS/EMHSS variance strata, groups of 18 adjacent main study variance strata will be collapsed. Thus, a total of 50 MHSS/EMHSS variance strata will be formed. The variance replicates will be retained from the main study.

Exhibit 3 MHSS/EMHSS Sample Weight Components

Dwelling Unit-Level Weight Components	
#1.	Inverse Probability of Selecting Census Tract
#2.	Inverse Probability of Selecting Segment
#3.	Quarter Segment Weight Adjustment
#4.	Subsegmentation Inflation Adjustment
#5.	Inverse Probability of Selecting Dwelling Unit
#6.	Inverse Probability of Added/Subsampled Dwelling Unit
#7.	Dwelling Unit Release Adjustment
#8.	Dwelling Unit Nonresponse Adjustment
#9.	Dwelling Unit Poststratification Adjustment
#10.	Dwelling Unit Extreme Weight Adjustment

Person-Level Weight Components	
#11.	Inverse Probability of Selecting a Person within a Dwelling Unit
#12.	Selected Person Poststratification to Roster Adjustment
#13.	Person-Level Nonresponse Adjustment
#14.	Person-Level Poststratification Adjustment
#15.	Person-Level Extreme Weight Adjustment

MHSS/EMHSS Weight Components	
#16.	Inverse Probability of Selecting a Person for the MHSS/EMHSS
#17.	Person-Level Nonresponse Adjustment
#18.	Person-Level Poststratification Adjustment

5. General Sample Procedures for the Questionnaire Field Test

A Questionnaire Field Test (QFT) was conducted during the 2012 National Survey on Drug Use and Health (NSDUH). The field test overlapped with quarters 3 and 4 of the main study with data collection beginning on September 1, 2012, and ending on November 3, 2012. Approximately 2,000 persons were interviewed in order to test a redesigned questionnaire protocol for implementation in the 2015 NSDUH. The QFT was designed to address the impact that the redesign protocol had on the reporting of core substance use or the reporting on noncore survey items, on screening and interview timing, and on data quality. In addition, the QFT attempted to measure the feasibility of the redesign protocol based on field interviewer (FI) debriefings or respondent feedback to interviews. Data from the QFT were compared with data from the 2011 NSDUH and with data from quarters 3 and 4 of the 2012 main study. This chapter provides the sampling procedures for the QFT.

5.1 Target Population

Similar to NSDUH, the respondent universe for the QFT was the civilian, noninstitutionalized population aged 12 or older. In order to control costs, persons residing in Alaska and Hawaii as well as persons who do not speak English were excluded from the QFT. Therefore, the sample was representative of the English-speaking, noninstitutional population aged 12 or older in the contiguous United States.

5.2 Selection of State Sampling Regions and Segments

An estimated 213 segments were needed to yield approximately 2,000 completed interviews. To achieve national representation of the English-speaking population, a probability proportional to size sample of 213 (of 876) State sampling (SS) regions was selected. Alaska and Hawaii were excluded from the QFT sample to control costs. This design maximized the efficiency (i.e., increased precision) of the QFT estimates by reducing variation in the weights. In addition, this design had the benefit of placing the sample in heavily populated areas where a sufficient mix of FIs with various experience levels to meet staffing goals were expected. As shown in Table 5.1, a large portion of the sample was selected from the eight largest States.

Within each selected SS region, a sample of dwelling units (DUs) was drawn from the segment that was retired from use in quarter 1 of the 2012 NSDUH. If an insufficient number of eligible DUs remained in a segment, or if significant access problems were expected in a segment, the segment was replaced with the 2011 quarter 4 retired segment in the same SS region. A total of 6 segments were replaced because they had fewer than 10 eligible DUs remaining, and a total of 7 segments were replaced because of anticipated access problems in the segments.

Table 5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State

State	Population Rank (12 or Older)	Current Design	NSDUH SS Regions	Number of QFT SS Regions/ Segments (PPS)	2012 QFT Respondents
CA	1	3,600	48	23	170
TX	2	3,600	48	14	146
NY	3	3,600	48	11	105
FL	4	3,600	48	18	169
IL	5	3,600	48	10	72
PA	6	3,600	48	10	121
OH	7	3,600	48	7	103
MI	8	3,600	48	9	86
GA	9	900	12	6	60
NC	10	900	12	5	50
NJ	11	900	12	6	52
VA	12	900	12	6	83
MA	13	900	12	4	33
WA	14	900	12	5	46
IN	15	900	12	6	63
AZ	16	900	12	4	14
TN	17	900	12	4	51
MO	18	900	12	2	16
WI	19	900	12	4	38
MD	20	900	12	3	32
MN	21	900	12	4	36
CO	22	900	12	6	33
AL	23	900	12	4	45
SC	24	900	12	3	31
KY	25	900	12	3	28
LA	26	900	12	5	66
OR	27	900	12	1	8
OK	28	900	12	5	40
CT	29	900	12	5	41
IA	30	900	12	0	0
MS	31	900	12	0	0
AR	32	900	12	0	0
KS	33	900	12	2	19
NV	34	900	12	2	33
UT	35	900	12	6	63

(continued)

Table 5.1 Number of 2012 Questionnaire Field Test State Sampling Regions and Achieved Sample Sizes, by State (continued)

State	Population Rank (12 or Older)	Current Design	NSDUH SS Regions	Number of QFT SS Regions/ Segments (PPS)	2012 QFT Respondents
NM	36	900	12	1	4
WV	37	900	12	2	23
NE	38	900	12	3	25
ID	39	900	12	0	0
ME	40	900	12	2	12
NH	41	900	12	1	11
HI	42	900	12	0	0
RI	43	900	12	0	0
MT	44	900	12	1	16
DE	45	900	12	0	0
SD	46	900	12	0	0
AK	47	900	12	0	0
VT	48	900	12	0	0
ND	49	900	12	0	0
DC	50	900	12	0	0
WY	51	900	12	0	0
Total		**67,500**	**900**	**213**	**2,044**

SS = State sampling; PPS = probability proportional to size; QFT = Questionnaire Field Test.

5.3 Selection of Dwelling Units

The procedures described in Sections 3.3 and 3.4 were used to determine DU sample sizes and person probabilities of selection for the QFT. A sufficient number of DUs was drawn to account for the lower sample yield resulting from conducting interviews in English only. Data from the 2011 NSDUH were used to predict eligibility and nonresponse rates by State, and these estimates were used in the QFT sample allocation. The number of sample DUs was limited to 50 per segment in order to ensure a reasonable spread across QFT segments. DUs that were not selected for the 2010, 2011, and 2012 main studies were eligible for selection in the QFT.

Similar to the main study, a small reserve sample (20 percent) of DUs from each segment was selected, and the total sample was partitioned into four equal probability subsamples: 105/120, 5/120, 5/120, and 5/120. Although the majority of the sample (105/120) was released at the beginning of the QFT data collection period, having the additional sample partitions allowed for greater flexibility in controlling the sample size and provided the ability to ensure data collection goals were attained within the field period. Two additional 5 percent partitions (5/120) were released in all but six States[28] after 4 weeks of data collection.

[28] Additional sample was not released in the following States: Connecticut, New Mexico, Oregon, South Carolina, Utah, and Virginia.

A total of 5,358 DUs were sampled and yielded 2,044 completed interviews, as shown in Table 5.2. The half-open interval procedure for missed DUs (see Section 3.7) was implemented during the QFT, but it is not scheduled to be implemented in the 2014 or 2015 NSDUHs.

Table 5.2 Summary of the 2012 Questionnaire Field Test Sample Results

Statistic	Total	Rate
State Sampling (SS) Regions	213	
Segments	213	
Selected Dwelling Units	5,358	
Eligible Dwelling Units	4,623	0.86
Completed Screening Interviews	3,837	0.83
Selected Persons	2,823	
Eligible Persons (English Speaking)	2,760	0.98
Completed Interviews	2,044	0.74

5.4 Age Group Allocations

The respondent sample was allocated to the three major age groups in the following proportions: 25 percent aged 12 to 17, 25 percent aged 18 to 25, and 50 percent aged 26 or older. Within the 26 or older age group, 15 percent of the sample was allocated to persons aged 26 to 34, 20 percent of the sample was allocated to persons aged 35 to 49, and 15 percent was allocated to persons aged 50 or older. This sample allocation matched the planned allocation for the 2015 NSDUH partial redesign. One implication of the respondent sample allocation by age groups is a potential impact on QFT response rates. Retaining more of the 26 or older adults identified in households to complete interviews may have had a negative effect on unweighted interview response rates.[29] Specifically, respondents aged 50 or older have shown lower propensities to participate than other age groups in the NSDUH sample. The unweighted interview response rate for the QFT sample was 72.41 percent compared with 76.52 percent for the 2011 main study comparison sample and 79.31 percent for the 2012 quarters 3 and 4 main study comparison sample.[30] On the other hand, although a smaller proportion of 12 to 17 year olds were selected, this age group continued to drive the number of DUs needed (i.e., the age group continued to be sampled at the highest rate). Thus, fewer DUs were needed to yield the desired sample than would be needed under the current sample design.

5.5 Selection of Persons

After DUs were selected within each QFT segment, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. This roster information was used to select 0, 1, or 2 persons for the survey. Sampling rates were preset by segment and age group. Roster information was entered directly into the electronic screening instrument that automatically implemented this stage of selection based on the segment and age group sampling parameters. As illustrated in Table 5.2, a total of 2,823 people were selected from within 4,623 screened and eligible DUs, which yielded 2,044 completed interviews.

[29] Weighted interview response rates were not expected to change as a result of the age group allocation.

[30] Other factors influencing the interview response rates were the shorter data collection period and a limited number of FIs available for use in converting QFT refusals.

5.6 Creation of Variance Estimation Strata and Replicates

The nature of the stratified, clustered sampling design of the NSDUH main study and QFT samples requires that the design structure be taken into consideration when computing variances of survey estimates. Key nesting variables were created for the QFT and main study comparison samples to capture explicit stratification and to identify clustering.

To allow for comparisons between the QFT and main study samples, a common set of stratification and clustering variables were defined. Because SS regions serve as strata for the main study samples and as primary sampling units (PSUs) for the QFT sample, there was no direct way of capturing the covariance between the samples and using the entire main study sample. Instead, the approach used for the 1999 paper-and-pencil interviewing (PAPI) and computer-assisted interviewing (CAI) mode analysis was followed in developing a design structure that could be used to simultaneously analyze all three samples (Chromy, Davis, Packer, & Gfroerer, 2002, pp. 141-142). Steps in the process were as follows:

- Within QFT sampling strata (census regions), variance strata were generally formed by assigning two sequential QFT-selected SS regions to the same variance strata on the sorted sampling frame. Each sampled SS region was then assigned to a replicate (1 or 2). However, there were three QFT SS regions per variance strata for three randomly selected strata. This was necessary because there were an odd number of QFT SS regions selected in three of the census regions. Within these three strata, the third SS region was randomly assigned to either replicate 1 or replicate 2. This led to a total of 105 QFT variance strata, with two replicates per strata.
- Using the sorted QFT sampling frame of SS regions, the main study SS regions not selected for the QFT were assigned to QFT sampling strata sequentially, in accordance with the assignments of selected QFT SS regions. These assignments kept the number of SS regions per strata as equal as possible given the distribution of QFT sampled SS regions within the sorted SS region frame. For SS regions not selected for the QFT sample, the original replicate assignments of either replicate 1 or replicate 2 were maintained.

Although this approach to design structure variables does not fit the main study perfectly, it does capture the total variance and allows for taking advantage of any covariance induced by the overlapping SS regions between the samples.

References

American Psychiatric Publishing, Inc. (2000). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision* (DSM-IV-TR; doi: 10.1176/appi.books.9780890423349). Retrieved January 29, 2013, from http://www.psychiatryonline.com/resourceTOC.aspx?resourceID=1

Chen, P., Dai, L., Cribb, D., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (in press). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C, Phase I, Deliverable No. 39, RTI/ 0212800.001.107.004). Research Triangle Park, NC: RTI International.

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association.

Chromy, J. R. (1981). Variance estimators for a sequential sample selection procedure. In D. Krewski, R. Platek, & J. N. K. Rao (Eds.), *Current topics in survey sampling: Proceedings of the symposium held May 7-9, 1980, sponsored by the Ottawa Chapter and the Survey Research Methods Section of the American Statistical Association* (pp. 329-347). New York, NY: Academic Press.

Chromy, J. R. (1987). Design optimization with multiple objectives. In *Proceedings of the 1987 American Statistical Association, Survey Research Methods Section, San Francisco, CA* (pp. 194-199). Alexandria, VA: American Statistical Association.

Chromy, J., Davis, T., Packer, L., & Gfroerer, J. (2002). Mode effects on substance use measures: Comparison of 1999 CAI and PAPI data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 135-159). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Chromy, J. R., & Penne, M. A. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY* (pp. 552-554). Alexandria, VA: American Statistical Association.

Cochran, W. G. (1977). *Sampling techniques* (3rd ed.). New York, NY: John Wiley & Sons.

First, M. B., Spitzer, R. L, Gibbon, M., & Williams, J. B. W. (2002, November). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*, 184-189.

Kish, L. (1965). *Survey sampling*. New York, NY: John Wiley & Sons.

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research, 19*(S1), 49-60. doi: 10.1002/mpr.313.

Office of Management and Budget. (2003, June 6). *OMB Bulletin No. 03-04: Revised definitions of metropolitan statistical areas, new definitions of micropolitan statistical areas and combined statistical areas, and guidance on uses of the statistical definitions of these areas*. Retrieved January 29, 2013, from http://www.whitehouse.gov/omb/bulletins_b03-04

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*, 110-123. doi:10.1002/mpr.61.

RTI International. (2008). *SUDAAN®, Release 10.0* [computer software]. Research Triangle Park, NC: Author.

RTI International. (2011). *2012 National Survey on Drug Use and Health: Counting and listing general manual* (RTI/0212800, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C). Research Triangle Park, NC: Author.

U.S. Census Bureau, Redistricting Data Office. (2009). Appendix A: Geographic terms and concepts. In *2008 Redistricting data prototype (Public Law 94-171) summary file*. Washington, DC: Author. [Available as a PDF at http://www.census.gov/rdo/; see http://www.census.gov/rdo/pdf/pl94-171.pdf]

Appendix A:
2005 through 2013 NSDUH State Sampling Regions

2005-2013 NSDUH State Sampling Regions: Alabama

2005-2013 NSDUH State Sampling Regions: Alaska

2005-2013 NSDUH State Sampling Regions: Arizona

2005-2013 NSDUH State Sampling Regions: Arkansas

2005-2013 NSDUH State Sampling Regions: California

2005-2013 NSDUH State Sampling Regions: Colorado

2005-2013 NSDUH State Sampling Regions: Connecticut

2005-2013 NSDUH State Sampling Regions: Delaware

2005-2013 NSDUH State Sampling Regions: District of Columbia

2005-2013 NSDUH State Sampling Regions: Florida

2005-2013 NSDUH State Sampling Regions: Georgia

2005-2013 NSDUH State Sampling Regions: Hawaii

2005-2013 NSDUH State Sampling Regions: Idaho

2005-2013 NSDUH State Sampling Regions: Illinois

2005-2013 NSDUH State Sampling Regions: Indiana

2005-2013 NSDUH State Sampling Regions: Iowa

2005-2013 NSDUH State Sampling Regions: Kansas

2005-2013 NSDUH State Sampling Regions: Kentucky

2005-2013 NSDUH State Sampling Regions: Louisiana

2005-2013 NSDUH State Sampling Regions: Maine

2005-2013 NSDUH State Sampling Regions: Maryland

2005-2013 NSDUH State Sampling Regions: Massachusetts

2005-2013 NSDUH State Sampling Regions: Michigan

2005-2013 NSDUH State Sampling Regions: Minnesota

2005-2013 NSDUH State Sampling Regions: Mississippi

2005-2013 NSDUH State Sampling Regions: Missouri

2005-2013 NSDUH State Sampling Regions: Montana

2005-2013 NSDUH State Sampling Regions: Nebraska

2005-2013 NSDUH State Sampling Regions: Nevada

2005-2013 NSDUH State Sampling Regions: New Hampshire

2005-2013 NSDUH State Sampling Regions: New Jersey

2005-2013 NSDUH State Sampling Regions: New Mexico

2005-2013 NSDUH State Sampling Regions: New York

2005-2013 NSDUH State Sampling Regions: North Carolina

2005-2013 NSDUH State Sampling Regions: North Dakota

2005-2013 NSDUH State Sampling Regions: Ohio

2005-2013 NSDUH State Sampling Regions: Oklahoma

2005-2013 NSDUH State Sampling Regions: Oregon

2005-2013 NSDUH State Sampling Regions: Pennsylvania

2005-2013 NSDUH State Sampling Regions: Rhode Island

2005-2013 NSDUH State Sampling Regions: South Carolina

2005-2013 NSDUH State Sampling Regions: South Dakota

2005-2013 NSDUH State Sampling Regions: Tennessee

2005-2013 NSDUH State Sampling Regions: Texas

2005-2013 NSDUH State Sampling Regions: Utah

2005-2013 NSDUH State Sampling Regions: Vermont

2005-2013 NSDUH State Sampling Regions: Virginia

2005-2013 NSDUH State Sampling Regions: Washington

2005-2013 NSDUH State Sampling Regions: West Virginia

2005-2013 NSDUH State Sampling Regions: Wisconsin

2005-2013 NSDUH State Sampling Regions: Wyoming

Appendix B: 2012 NSDUH Procedure for Subsegmenting

B.1 Introduction

Subsegmenting is a statistical process used in the National Survey on Drug Use and Health (NSDUH) to reduce the size of the sample, which reduces the time and cost spent in the field for counting and listing. The precise and accurate application of subsegmenting procedures is most feasible when boundaries of subsegments can be formed using actual surface features, such as streets, rivers, and railroads. When such features cannot be used, listing the entire area segment is considered. Because subsegmenting is a sampling function, it must be carried out with the same high degree of scientific precision exercised in the other stages of sample development.

B.2 Determining Subsegmenting In-House

Prior to sending segments to the field for listing, segments that are candidates for in-house subsegmenting are identified based on the number of blocks, square miles, and dwelling units (DUs).[31] The two criteria for identifying candidate segments are as follows:

- number of blocks > 1, square miles ≥ 75, and DU count ≥ 150; or
- number of blocks > 1 and DU count ≥ 300.

Candidate segments then are evaluated to determine whether they can be subsegmented without input from the field. If feasible, the subsegmenting is performed in-house prior to sending the segment to the field for listing. This step expedites the process and saves time and field expenses for very large segments.

B.3 Determining Subsegmenting While in the Field

If a certified lister is counting a segment and determines that the DU count is greater than 400, the segment is too large and must be subsegmented. The lister then mails the segment materials back to the sampling support office. When the segment is in-house, standard subsegmenting procedures are followed using the street segment counts obtained by the lister.

In the field, some of the segments that were originally subsegmented in-house (as described in Section B.2) may still be too large to list. Additional subsegmenting is required for one of the following reasons: (1) the area experienced high growth, and the census counts used in the initial subsegment were outdated, or (2) there was not enough information available during the first subsegment, and the initial subsegment was still too large to list. In the latter case, the initial subsegment was done to make the counting more manageable, but a second subsegment had to be done to make listing feasible. The initial subsegment then is counted by the lister and sent back to the sampling support office where standard subsegmenting procedures are applied.

[31] DU counts were obtained from 2000 census data supplemented with revised population counts from Claritas.

B.4 Standard Subsegmenting Procedures

Once it is determined that subsegmenting is required, the following procedure is used:

Step 1: On the basis of the count, the segment is divided into areas (list units) containing no fewer than 100 DUs. If available, actual surface features are used to form new boundaries between divisions. An attempt to maintain balance between divisions is made (the largest list unit should not contain more than 1½ times the number of DUs contained in the smallest unit).

Step 2: After properly dividing the segment into list units, the units are lettered consecutively with capital letters (A, B, C, ...) starting with the list unit and including the northeast (or most appropriate) corner of the segment and continuing clockwise around the segment.

Step 3: Using a subsegmenting worksheet, one of the list units is randomly selected to be listed. On the worksheet, the number of DUs in each list unit is recorded and accumulated. A random number generated for each segment is multiplied by the total accumulated DUs. The product then is rounded up, and the list unit whose cumulative DUs is greater than or equal to the product is selected for listing.

After the segment materials have been returned to the field, only the selected unit is listed. All counts used in the subsegmenting process are retained so that weights can be adjusted to reflect the entire area segment.

Appendix C: 2012 NSDUH Procedure for Adding Missed Dwelling Units

C.1 Introduction

The 2012 National Survey on Drug Use and Health (NSDUH) requires field interviewers (FIs) to visit sample segments and screen and interview dwelling units (DUs) that were selected from an ordered list. The list of DUs, which includes housing units (HUs) and group quarters (GQs), was constructed by the counting and listing staff during the summer and fall of 2011. Because the listing was done a short time before the 2012 screening and interviewing activities began, no major discrepancies were expected. However, factors such as new construction, demolition, and inaccurate listing may be present in some cases. More commonly, DUs may have been "hidden" and therefore overlooked by the counter and lister.

For all DUs to be given a chance of being selected, NSDUH has a procedure for locating and adding missed DUs. It requires FIs to look on the property of selected DUs and between that DU and the next listed DU (i.e., the half-open interval rule, as described in Section 3.7 of Chapter 3). In 2000, the rule was modified such that the half-open interval is closed on each map page. Therefore, if the selected DU is the last on a page, the "next listed DU" will be the first one listed on the same page. If the number of added DUs linked to any particular DU does not exceed 5 or if the number for the entire segment is less than or equal to 10, the FI is instructed to consider these DUs as part of his or her assignment. However, if either of these limits is exceeded, the FI will contact the sampling support office for subsampling to be considered.

This appendix outlines the procedures for sampling staff to use when discrepant segments are found in the field. For this appendix, procedures for adding missed DUs are classified into three categories: adding HUs, adding GQ units, and "busts."

C.2 Motivation

Prior to the 1999 survey, if the number of added DUs exceeded the defined limits, the added DUs were subsampled at the same rate of the original selection for the segment. To maintain unequal weighting effect and to control costs associated with adding DUs, a new subsampling procedure was implemented:

Number of Added DUs	Sampling Rate
0	No action
1 to 10	Automatic (all DUs added to the sample)
11 to 25	1/2
26 to 40	1/3
41 to 50	1/4
50 or more	1/5

C.3 Procedure for Adding Housing Units

This section refers to HUs that are obtained through the half-open interval rule. This method of dealing with added HUs is preferable to all others because it is probability-based and maintains the integrity of the sample. When possible, this methodology will be used to resolve added DU problems.

1. Once the limit of 5 (or 10) rule is exceeded, the FI should stop screening and interviewing activities on added HUs and contact the sampling support office. The FI will be instructed to do a quick check of the segment to see if any other listing problems might arise. At this time, the FI will complete a paper list of added HUs for the entire segment.
2. Once the final list of added HUs has been received by the sampling support office, the following is to be done:
   (a) Sampling will examine the added HUs and determine whether they are linked to a sample dwelling unit (SDU).
   (b) If the number of added HUs linked to any *one* SDU exceeds 50, these units will be treated as a "bust" (see Section C.6).
   (c) If the number of added HUs linked to any *one* nonsampled DU exceeds 50, these units also will be treated using the procedure for "busts" (see Section C.6).
   (d) Sampling staff will calculate the total number of added DUs by adding the number of sampling units obtained through the "bust" procedure to the number of added DUs obtained through the half-open interval rule.
   (e) If the total number of added DUs exceeds 10, a subsampling rate will be determined using the criteria above.
3. The computing division will add the DUs to the system and subsample if necessary:
   (a) Data entry of the added DUs will be done. Entries will be made for all units that collectively qualify as a "bust" and units obtained through the half-open interval rule—not for all missed DUs found in the segment. The link number then will be entered and a line number will be assigned.[32] For DUs obtained through the "bust" procedure, the sampling link number (SLN) also will be recorded. Finally, it will be necessary to check that none of the DUs has already been entered in the iPAQ (i.e., the handheld computer) so that DUs do not appear in the system twice.
   (b) DUs will be selected from the added DUs at the rate defined above. The subsampling rate will be recorded in a data field.
   (c) Probabilities of selection will be brought over as appropriate for the segment.
   (d) A random number will be added for the iPAQ selection algorithm.

[32] During the listing process, each DU is written on a separate line on the listing form and assigned a corresponding line number (i.e., the number of lines equals the number of DUs). The added DUs are assigned the next available line number.

4. Selected DUs will be added to the FI's assignment during the next transmission.
5. A sample weight will be assigned to each added DU. If the total number of added DUs is less than or equal to 10, each added DU is assigned the weight of the original selected DUs in the segment. If subsampling is required, the selected DU weight is adjusted by the inverse of the subsampling rate for each added DU.

C.4 Procedure for Adding Group Quarters Structures

In the case of an entire GQ structure not being listed (or erroneously being listed as an HU), the half-open interval rule will be applied. For example, if the DU preceding the GQ was selected, or if the HU that is really a GQ was selected, the entire GQ structure will be added to the sample. The exception to this rule will be if the number of GQ units in the missed GQ structure exceeds 50. In this last case, the "bust" procedure will be applied (see Section C.6).

C.5 Procedure for Adding Group Quarters Units

In the case of discrepant GQ listings, the number of sampling units (rooms, persons, or beds) and the number of selected units will be known in advance. If the actual number of sampling units equals the amount listed in advance, the iPAQ will only need to be notified of the new unit type in order to function properly. However, if the actual units do not equal the advance units, two approaches will be taken.

C.5.1 Number of Actual GQ Units Less Than Number of Advance GQ Units

In the case that there are extra GQ units listed, the units at the end of the list will be assigned an ineligible code, such as "Listing Error." All other units will remain eligible.

C.5.2 Number of Actual GQ Units Greater Than Number of Advance GQ Units

If there are more GQ units in the structure than were previously listed, a complete list will be made, and the units will be consecutively numbered. Assume, for example, that 11 units were listed and 45 were actually found. Also, assume that units 1, 5, and 10 were selected for screening and interviewing (indicated in bold).

Original list:	**1**
	2
	3
	4
	5
	6
	7
	8
	9
	10
	11

The additional units then will be numbered consecutively, and an SLN corresponding to each of the originally listed units will be assigned. Next, the added GQ units with SLNs corresponding to the original selected units will be added to the sample:

Unit Number	SLN
12	**1**
13	2
14	3
15	4
16	**5**
17	6
18	7
19	8
20	9
21	**10**
22	11
23	**1**
24	2
25	3
26	4
27	**5**
28	6
29	7
30	8
31	9
32	**10**
33	11
34	**1**
35	2
36	3
37	4
38	**5**
39	6
40	7
41	8
42	9
43	**10**
44	11
45	**1**

C.6 "Busts"

A "bust" is any segment listing with a major discrepancy (defined by 150 or more total unlisted units or 50 or more added DUs linked to any one SDU) or that is completely unrepresentative of what is actually found. In the case of a fictitious listing, a lister will relist the segment as quickly as possible. Otherwise, the following approach will be employed.

First, if any DUs have disappeared since the time of the listing, all selected "disappears" will be assigned an "ineligible" final screening code. Then any new DUs will be listed consecutively, assigned an SLN, and added to the sample if the SLN corresponds to the line number of an originally selected DU. Note that if the DU was coded as ineligible in the first step, the new DUs having its line number as the SLN still will be added. This procedure is identical to the procedure for adding extra GQ units; however, the list can contain any combination of HUs and GQ units in this case. Again, if the number of DUs added is greater than 10, then resampling will occur from all nonfinalized DUs as in Section C.3.

C.7 Quality Control

To ensure quality, the sampling support office will employ several quality control checks:

- Sampling staff will ensure that the correct information has been keyed by data entry.
- Checks within the computing division will be performed.
- Sampling staff will check the number of selected DUs and the person probabilities of selection assigned to each DU selected in the subsampling routine.

Instrumentation Protocol Changes

The 2012 National Survey on Drug Use and Health (NSDUH) instrumentation protocols underwent several modifications from the protocols employed for the 2011 NSDUH. A summary of the changes for the 2012 NSDUH can be divided into three topics: (1) the computer-assisted interviewing (CAI) screening instrument, (2) the CAI questionnaire, and (3) the interviewer materials.

1. Screening Instrument

Using the 2011 Screening Application as a base, no changes were made to the 2012 Screening Application other than updating dates.

2. CAI Questionnaire

Using the 2011 CAI Questionnaire as a base, questions and routing were modified for the 2012 CAI Questionnaire. Listed below are the CAI Questionnaire changes for 2012:

- The CAI instrument version was updated in the Introduction section.
- CGREF5, which is asked of respondents who refuse to answer an initial item about smoking tobacco in a pipe in the tobacco module, was edited. As opposed to asking respondents if they have ever smoked a pipe, the question now asks if they have ever smoked *tobacco in* a pipe.
- Added new morphine dosages to Box 11 of Pillcard A that accompanies questions in the Pain Relievers module.
- State program names for Medicaid, CHIP and TANF in the Health Insurance and Income modules were updated.

3. Interviewer Materials

The 2012 interviewer materials and protocols for using those materials were based on the 2011 interviewer materials and protocols. In addition to minor wording revisions and year updates (2011 to 2012), the following changes were made for the 2012 interviewer materials:

- The Quality Control Forms and Reference Date Calendars were printed on peach-colored paper in 2012.
- The shading in the month and date fields [MM/DD] in the interviewer portion of the Quality Control Form was lightened to make it easier to read information written in these fields.
- The Introduction and Informed Consent for Interview Respondents Age 12-17 [in the Showcard Booklet] was updated to clarify the FI instructions shown above the parental permission script at the top of the page. This text was

revised to read, “FIRST, READ THE PARENTAL PERMISSION SCRIPT BELOW AND OBTAIN PERMISSION FROM THE PARENT.”

- The RTI/SAMHSA Fact Sheet was updated with more current information on SAMHSA and RTI.
- The Q&A (Question and Answer) Brochure included an updated list of primary objectives for NSDUH to be consistent with other NSDUH materials. The text under the "How is the Study Administered?” heading was updated to state the interview takes “about an hour.”
- The NSDUH Respondent Web site content included updated links to information on NSDUH. The NSDUH in the News web page was updated with links to more recent articles and reports and a video containing footage from the most recent NSDUH press conference was included. The listing of primary objectives for NSDUH was updated to be consistent with other NSDUH materials
- *The NSDUH Report: Illicit Drug Use Among Older Adults* (Center for Behavioral Health Statistics and Quality, 2011) and two Data Spotlights: *OAS Data Spotlight: 12 Year Olds More Likely to Use Inhalants Than Cigarettes or Marijuana* (Office of Applied Studies, 2010) and *Data Spotlight: Young Alcohol Users Often Get Alcohol from Family or Home* (Center for Behavioral Health Statistics and Quality, 2011), were added to the interviewer materials for use with respondents.
- The Showcard Booklet underwent a few updates for 2012:
  - Additional dosages of ER and IR morphine were added to Pillcard A and the survey year was updated on bottom of all pillcards.
  - An Informed Consent Reference Guide was added to the Job Aids section for easy reference
  - The Steps to Maximizing Data Quality exhibit in the Job Aids section was updated to include a reminder that each protocol listed must be followed at all times. Also, the FI Manual references and text in the exhibit was updated to reflect current information.
  - The Interview Process exhibit in the Job Aids section was updated to include reminders on plugging in the headphones as part of the computer set-up and working on Data Quality Knowledge Notebook entries while the respondent is completing the ACASI portion of the interview.
- The Appointment Card was updated to simplify the text and re-formatted to present a more professional layout.
- The Sorry I Missed You Card was re-formatted to present a more professional layout.
- The Interview Payment Receipt was updated to reflect current information on SAMHSA’s Treatment Referral Helpline.

- The Newspaper Article handout was updated with two new articles featuring information on Methamphetamine and marijuana use.
- The NSDUH Highlights were updated to reflect the 2010 study results.
- The Doorperson Letter was updated to include Joel Kennet's name and signature. The dates in the FI ID badge were updated and the letterhead was edited to make it similar to the Lead Letter.
- The SAMHSA Authorization letter was updated with the current year and contract number.
- The Data Collection Agreement was updated to include stronger language regarding the consequences and penalties for falsification of any NSDUH data.
- The ePTE Working Copy was updated with the 2012 project number.

Copies of the 2012 interviewer materials can be found in Section 7, Written Communication, in the *2012 NSDUH Methodological Resource Book.*

References

Center for Behavioral Health Statistics and Quality. (2011, September). Illicit Drug Use Among Older Adults. *The NSDUH Report*. [Available at http://www.samhsa.gov/data/Short_Reports.aspx]

Center for Behavioral Health Statistics and Quality. (2011, February). Young Alcohol Users Often Get Alcohol from Family or Home. *Data Spotlight*. [Available at http://www.samhsa.gov/data/Spotlights.aspx]

Office of Applied Studies. (2010, March). 12 Year Olds More Likely to Use Inhalants Than Cigarettes or Marijuana. *OAS Data Spotlight*. [Available at http://www.samhsa.gov/data/Spotlights.aspx]

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

DATA COLLECTION FINAL REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800
Deliverable No. 41

Authors:

Christine Clark
Lee Ellen Coffey
David C. Hunter
Ilona Johnson
Shuangquan Liu
Megan C. Livengood
Lauren M. Machingo
Peilan C. Martin
Allison C. McKamey
Katherine B. Morton
Susan K. Myers
James Pratt
Beth H. Riggsbee
Dustin Williams
Hilary Zelko

Project Director: Thomas G. Virag

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

September 2013

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

DATA COLLECTION FINAL REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800
Deliverable No. 41

Authors:

Christine Clark
Lee Ellen Coffey
David C. Hunter
Ilona Johnson
Shuangquan Liu
Megan C. Livengood
Lauren M. Machingo
Peilan C. Martin
Allison C. McKamey
Katherine B. Morton
Susan K. Myers
James Pratt
Beth H. Riggsbee
Dustin Williams
Hilary Zelko

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

September 2013

Table of Contents

Chapter **Page**

List of Tables vii
List of Exhibits xi

1. Introduction 1

2. Sampling and Counting and Listing Operations 3
   2.1 Overview of Sampling Procedures 3
   2.2 Recruiting and Training for Counting and Listing 4
   2.3 Counting and Listing Procedures 5
   2.4 Added Dwelling Units 6
   2.5 Problems Encountered 7
      2.5.1 Controlled Access 7
      2.5.2 Segments with Reassigned Quarters 8
      2.5.3 Edited Addresses 8
   References 14

3. Data Collection Staffing 15
   3.1 Regional Directors 15
   3.2 Regional Supervisors 15
   3.3 Field Supervisors 16
   3.4 Field Interviewers and Traveling Field Interviewers 16
   3.5 Problems Encountered 19
      3.5.1 Continued Staffing Shortfall in Certain Areas 19
      3.5.2 FI Turnover 19

4. Preparation of Survey Materials 33
   4.1 Electronic Screening 33
   4.2 Questionnaire Development 33
      4.2.1 CAI Instrument 33
      4.2.2 Spanish Translations 33
   4.3 Manuals and Miscellaneous Materials Development 33
      4.3.1 Manuals 33
      4.3.2 Miscellaneous Materials 35
   4.4 Submission of the 2012 NSDUH IRB Package 36
   4.5 Preparation for New-to-Project Field Interviewer Training 37
      4.5.1 Home Study Package 37
      4.5.2 New-to-Project Training Supplies 38
      4.5.3 New-to-Project Bilingual Training 40
   4.6 Preparation for Veteran Field Interviewer Training 40
      4.6.1 Veteran Training iLearning and Data Collection Preparations Packages 40
      4.6.2 Veteran Training Supplies 41

Table of Contents (continued)

Chapter **Page**

4.7 Preparation for Field Data Collection ... 42
4.7.1 Assignment Materials ... 43
4.7.2 Bulk Supplies ... 43
4.8 Website Development ... 43
4.8.1 Case Management System ... 43
4.8.2 NSDUH Respondent Website ... 44
4.9 Maintaining NSDUH Equipment ... 44
4.10 Problems Encountered ... 44
References ... 47

5. Field Staff Training ... 49
5.1 Management Training Programs ... 49
5.2 New-to-Project Field Interviewer Training Sessions ... 49
5.2.1 Design ... 49
5.2.2 Staffing ... 50
5.2.3 Content of New-to-Project Field Interviewer Training Sessions ... 52
5.2.4 New-to-Project Bilingual Training ... 54
5.2.5 Mentoring of New-to-Project Graduates ... 54
5.2.6 New-to-Project Post-Training iLearning ... 55
5.3 Veteran Field Interviewer Training Sessions ... 55
5.3.1 Design ... 55
5.3.2 Staffing ... 56
5.3.3 Train-the-Trainers ... 56
5.3.4 Content of Veteran Field Interviewer Training Sessions ... 56
5.3.5 Special Veteran Training Sessions ... 57
5.4 Ongoing Training ... 57
5.5 Periodic Evaluations ... 57
5.6 Problems Encountered ... 57

6. Data Collection ... 71
6.1 Contacting Dwelling Units ... 71
6.1.1 Lead Letter ... 71
6.1.2 Initial Approach ... 71
6.1.3 Introduction, Study Description, and Informed Consent ... 71
6.1.4 Callbacks ... 72
6.2 Dwelling Unit Screening ... 72
6.3 Within-Dwelling Unit Selection ... 72
6.4 Interview Administration ... 73
6.4.1 Informed Consent and Getting Started ... 73
6.4.2 Computer-Assisted Interviews ... 74
6.4.3 End of Interview Procedures ... 74
6.5 Data Collection Management ... 75
6.6 Controlled Access Procedures ... 76

Table of Contents (continued)

Chapter **Page**

6.7 Refusal Conversion Procedures 77
6.8 Problems Encountered 78
6.8.1 Size and Scope of the Project 78
6.8.2 Interviewing Staff Turnover 78
6.8.3 Refusals 78
6.8.4 Typical Data Collection Concerns 79
6.8.5 Adverse Weather 79
6.8.6 iPAQ 79
6.8.7 CAI and iPAQ Patches 79

7. Data Collection Results 115
7.1 Overview 115
7.2 Screening Response Rates 115
7.3 Interview Response Rates 115
7.4 Spanish Interviews 116
7.5 Interviewer Assessment of the Interview 116
7.6 Number of Visits 116

8. Quality Control 361
8.1 Field Supervisor and Interviewer Evaluation 361
8.1.1 Regular Conferences 361
8.1.2 New-to-Project Training and Training Evaluations 361
8.1.3 Veteran Training and Ongoing FI Knowledge Evaluations 362
8.1.4 Field Interviewer Observations 362
8.1.5 FS Evaluations of FIs 362
8.1.6 FI Exit Interviews 363
8.2 Data Quality Team 363
8.3 Data Quality Monitoring 364
8.3.1 Field Management Data Quality Reports 364
8.3.2 Data Quality Team Data Quality Reports 365
8.4 Verification of Completed Cases 366
8.4.1 In-House Verification 367
8.4.2 Field Verification 368
8.4.3 Verification Reporting Tools 369
8.5 Industry and Occupation Coding 370
Reference 399

Table of Contents (continued)

Appendices **Page**

A New-to-Project Home Study Cover Memo A-1

B New-to-Project Electronic Home Study Exercises B-1

C Veteran FI Training iLearning Courses Memo C-1

D Data Collection Preparations Memo D-1

E Verification Scripts E-1

F U.S. Census Bureau Industry and Occupation Coding Report F-1

List of Tables

Table		Page
1.1	Schedule of Major Data Collection Activities	2
2.1	2012 NSDUH Sampling Summary	9
2.2	2012 Segments with Added Dwelling Units	10
3.1	Distribution of 2012 Veteran Interviewers, by Race and Gender	21
3.2	Distribution of Interviewers Hired in 2012, by Race and Gender	21
3.3	Distribution of All 2012 Interviewers, by Race and Gender	21
3.4	Distribution of 2012 Veteran Interviewers, by Gender and Language Ability	21
3.5	Distribution of Interviewers Hired in 2012, by Gender and Language Ability	21
3.6	Distribution of All 2012 Interviewers, by Gender and Language Ability	22
5.1	2012 NSDUH FI Training Programs	59
5.2	2012 NSDUH Veteran Training Sites	60
5.3	Results from New-to-Project Post-Training iLearning	60
5.4	Results from Veteran Training iLearning Courses	60
5.5	Results from Periodic iLearning Evaluations	61
6.1	2012 NSDUH CAI Audit Trail Timing Data: Total Interview Time (Minutes) with FI Observation Section	80
6.2	2012 NSDUH CAI Audit Trail Timing Data: Introduction and Core Demographics Sections	81
6.3	2012 NSDUH CAI Audit Trail Timing Data: Total ACASI	82
6.4	2012 NSDUH CAI Audit Trail Timing Data: Tutorial Section	83
6.5	2012 NSDUH CAI Audit Trail Timing Data: Total Core Substances Sections	84
6.6	2012 NSDUH CAI Audit Trail Timing Data: Tobacco Section	85
6.7	2012 NSDUH CAI Audit Trail Timing Data: Alcohol Section	86
6.8	2012 NSDUH CAI Audit Trail Timing Data: Marijuana Section	87
6.9	2012 NSDUH CAI Audit Trail Timing Data: Cocaine and Crack Sections	88
6.10	2012 NSDUH CAI Audit Trail Timing Data: Heroin Section	89
6.11	2012 NSDUH CAI Audit Trail Timing Data: Hallucinogens Section	90
6.12	2012 NSDUH CAI Audit Trail Timing Data: Inhalants Section	91
6.13	2012 NSDUH CAI Audit Trail Timing Data: Total Pill Sections	92
6.14	2012 NSDUH CAI Audit Trail Timing Data: Total Noncore Sections	93
6.15	2012 NSDUH CAI Audit Trail Timing Data: Special Drugs Section	94

List of Tables (continued)

Table **Page**

6.16 2012 NSDUH CAI Audit Trail Timing Data: Risk/Availability Section 95

6.17 2012 NSDUH CAI Audit Trail Timing Data: Blunts Section 96

6.18 2012 NSDUH CAI Audit Trail Timing Data: Substance Dependence and Abuse Section..... 97

6.19 2012 NSDUH CAI Audit Trail Timing Data: Market Information for Marijuana Section..... 98

6.20 2012 NSDUH CAI Audit Trail Timing Data: Prior Substance Use Section 99

6.21 2012 NSDUH CAI Audit Trail Timing Data: Special Topics, Drug Treatment, and Health Care Sections 100

6.22 2011 NSDUH CAI Audit Trail Timing Data: Adult Mental Health Service Utilization Section 101

6.23 2012 NSDUH CAI Audit Trail Timing Data: Social Environment Section..... 102

6.24 2012 NSDUH CAI Audit Trail Timing Data: Parenting Experiences Section..... 103

6.25 2012 NSDUH CAI Audit Trail Timing Data: Youth Experiences Section 104

6.26 2012 NSDUH CAI Audit Trail Timing Data: Mental Health Section..... 105

6.27 2012 NSDUH CAI Audit Trail Timing Data: Adult Depression Section 106

6.28 2012 NSDUH CAI Audit Trail Timing Data: Youth Mental Health Service Utilization Section 107

6.29 2012 NSDUH CAI Audit Trail Timing Data: Adolescent Depression Section..... 108

6.30 2012 NSDUH CAI Audit Trail Timing Data: Consumption of Alcohol Section..... 109

6.31 2012 NSDUH CAI Audit Trail Timing Data: Total FI-Administered Back-End Demographics Section 110

6.32 2012 NSDUH CAI Audit Trail Timing Data: Back-End Demographics Section 111

6.33 2012 NSDUH CAI Audit Trail Timing Data: Income Section..... 112

6.34 2012 NSDUH CAI Audit Trail Timing Data: FI Observation Section 113

6.35 2012 NSDUH CAI Audit Trail Timing Data: Back-End Demographics Section among Persons Aged 15 or Older, by Employment Status..... 114

7.1 Summary of NSDUH Results 117

7.2 2012 Screening Results, by Population Density (Unweighted Percentages)..... 118

7.3 2012 Screening Results, by Population Density (Weighted Percentages)..... 119

7.4 2012 Screening Results, by Final Result and Population Density (Unweighted Percentages) 120

List of Tables (continued)

Table **Page**

7.5 2012 Screening Results, by Final Result and Population Density (Weighted Percentages) ... 121

7.6 2012 Screening Results—Completion Rate, by State and Population Density (Unweighted Percentages) ... 122

7.7 2012 Screening Results—Completion Rate, by State and Population Density (Weighted Percentages) ... 124

7.8 2012 Screening Results—Eligibility Rate, by State (Unweighted Percentages) ... 126

7.9 2012 Screening Results—Eligibility Rate, by State (Weighted Percentages) ... 128

7.10 2012 Screening Results—Completion Rate, by State (Unweighted Percentages) ... 130

7.11 2012 Screening Results—Completion Rate, by State (Weighted Percentages) ... 132

7.12 2012 Screening Results—Nonresponse Rate, by State (Unweighted Percentages) ... 134

7.13 2012 Screening Results—Nonresponse Rate, by State (Weighted Percentages) ... 136

7.14 2012 Screening Refusal Results (Total United States) (Unweighted Percentages) ... 138

7.15 2012 Screening Refusal Results (Total United States) (Weighted Percentages) ... 138

7.16 Interview Response Rates, by Age (Total United States) ... 165

7.17 Interview Response Rates, by Small Age Groups (Total United States) ... 165

7.18 2012 Interview Results, by Gender and Age (Unweighted Percentages) ... 166

7.19 2012 Interview Results, by Gender and Age (Weighted Percentages) ... 167

7.20 2012 Interview Results, by Age (Total United States) (Unweighted Percentages) ... 168

7.21 2012 Interview Results, by Age (Total United States) (Weighted Percentages) ... 168

7.21a 2012 Interview Results, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) ... 220

7.22 2012 Interview Results, by Age and Incomplete Interview Result (Total United States) (Unweighted Percentages) ... 224

7.23 2012 Interview Results, by Age and Incomplete Interview Result (Total United States) (Weighted Percentages) ... 224

7.23a 2012 Interview Results, by Age and Race/Ethnicity, and Incomplete Interview Result (Total United States) (Weighted Percentages) ... 276

7.24 2012 Interview Refusal Reasons, by Age (Total United States) (Unweighted Percentages) ... 281

7.25 2012 Interview Refusal Reasons, by Age (Total United States) (Weighted Percentages) ... 281

List of Tables (continued)

Table		Page
7.25a	2012 Interview Refusal Reasons, by Age and Race/Ethnicity (Total United States) (Weighted Percentages)	332
7.26	2012 Interview Results, by Small Age Groups and Gender (Unweighted Percentages)	336
7.27	2012 Interview Results, by Small Age Groups and Gender (Weighted Percentages)	339
7.28	2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Unweighted Percentages)	342
7.29	2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Weighted Percentages)	346
7.30	2012 Interview Results—Spanish Interviews, by State (Unweighted Percentages)	350
7.31	2012 Interview Results—Spanish Interviews, by State (Weighted Percentages)	352
7.32	2012 Interview Results—Spanish Interviews, by Age and Type of County (Unweighted Percentages)	354
7.33	2012 Interview Results—Spanish Interviews, by Age and Type of County (Weighted Percentages)	354
7.34	2012 English and Spanish Interviews Conducted, by Region	355
7.35	2012 English and Spanish Interviews Conducted, by Population Density	355
7.36	2012 Interviewer's Assessment of Respondent's Level of Understanding, by Age and Race/Ethnicity of Respondent	356
7.37	2012 Interviewer's Assessment of Respondent's Level of Cooperation during Interview, by Age and Race/Ethnicity of Respondent	357
7.38	2012 Interviewer's Assessment of Respondent's Level of Privacy during Interview, by Age and Race/Ethnicity of Respondent	358
7.39	2012 Interviewer's Assessment of How Often Respondent Revealed Answers in ACASI Sections, by Age and Race/Ethnicity of Respondent	359
7.40	Number of Visits Required To Complete Screening	360
7.41	Number of Visits Required To Complete Interview	360
8.1	2012 NSDUH FI Exit Interview Results	371
8.2	2012 NSDUH FI Exit Interviews—Most Important Reason for Resignation	377
8.3	2012 NSDUH Phone Verification Results—Noninterview Screening Cases	378
8.4	2012 NSDUH Phone Verification Results—Interview Cases	378
8.5	2012 NSDUH Field Verification Results—Noninterview Screening Cases	379

List of Tables (continued)

Table **Page**

8.6 2012 NSDUH Field Verification Results—Interview Cases 379
8.7 2012 NSDUH Field Verification Results—Field Interviewers 379

List of Exhibits

Exhibit **Page**

2.1 2012 NSDUH Sample Design Summary 11
3.1 NSDUH Management Chart 23
3.2 Data Collection Agreement 30
3.3 Flow of FI Recruiting Activity 31
4.1 2012 iPAQ Updates 45
4.2 2012 CAI Changes 46
5.1 Daily FI Training Evaluation 62
5.2 Mentoring Instructions 64
8.1 Steps to Maximize Data Quality 380
8.2 2012 NSDUH Field Interviewer Exit Interview 382
8.3 Overview of NSDUH Noninterview Screening Verification Process 387
8.4 Overview of NSDUH Interview Verification Process 388
8.5 Quality Control Form 389
8.6 CAI Mail Verification Letter 391
8.7 Short FI-Level Verification Report—Page 1 392
8.8 Short FI-Level Verification Report—Page 2 394
8.9 Short FI-Level Verification Report Problem Codes 396
8.10 Field Verification Summary Report 398

1. Introduction

The 2012 National Survey on Drug Use and Health (NSDUH) was the 32nd in a series of general population surveys designed to provide annual nationwide data on substance use patterns and behaviors in the United States. Continuing the expanded sample design first implemented in 1999, the scope of the 2012 survey allowed for the production of data estimates for the Nation, each of the 50 States, and the District of Columbia. Prior to 2002, the survey was known as the National Household Survey on Drug Abuse (NHSDA).[1]

NSDUH is funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), an agency of the U.S. Public Health Service in the U.S. Department of Health and Human Services. SAMHSA contracted with RTI International[2] to conduct activities including sampling, counting and listing, screening, interviewing, data processing, and reporting. This report examines the preparations and procedures used in carrying out the data collection tasks and also presents the results of data collection.

As an overview, data collection preparatory work on the 2012 NSDUH began in February 2011. Following a January 2012 training program for all returning veteran field interviewers, data collection work began on January 4, 2012, and was completed by December 20, 2012. The field staff of approximately 700 field interviewers worked to complete a total of 68,309 interviews using computer-assisted interviewing.

Table 1.1 provides approximate time frames for the various tasks completed.

The remainder of this report addresses the following topics relating to data collection for the 2012 NSDUH: Sampling and Counting and Listing Operations, Data Collection Staffing, Preparation of Survey Materials, Field Staff Training, Data Collection, Data Collection Results, and Quality Control.

[1] Throughout this report, a reference made to a past NSDUH implies a past NHSDA because the two names refer to the same annual survey.

[2] RTI International is a trade name of Research Triangle Institute.

Table 1.1 Schedule of Major Data Collection Activities

Activity	Approximate Time Frame
Conduct 2012 data collection preparations kickoff meeting.	February 2, 2011
Recruit listing staff.	March–April 2011
Conduct counting and listing and create lists of sample dwelling units.	April–December 2011
Prepare computerized screening and interviewing programs.	May–November 2011
Adjust 2011 management staff for 2012 because of new territory alignments.	Fall 2011
Recruit field interviewers for Quarter 1, 2012 (replacement staff also hired throughout the year as needed).	October–December 2011
Prepare manuals and materials for trainings.	May 2011–January 2012
Conduct veteran field interviewer training sessions.	January 2012
Conduct new-to-project field interviewer training sessions.	January–September 2012
Conduct and manage screening and interviewing operations.	January 4–December 20, 2012
Conduct verification operations.	January 12–December 28, 2012

2. Sampling and Counting and Listing Operations

2.1 Overview of Sampling Procedures

A coordinated 5-year sample design was developed for the 2005–2009 National Surveys on Drug Use and Health (NSDUHs). The 2010–2012 samples are extensions of the 5-year sample. The sample design for the 2012 NSDUH, as a subsample of the multiyear study, consisted of a deeply stratified, multistage area probability design. At the end of this chapter, Exhibit 2.1, in conjunction with Table 2.1, presents details of the sample design. The coordinated 2005–2009 design uses a 50 percent overlap in second-stage units (area segments) between each successive year of the 5-year study following completion of the 2005 survey. The 2012 NSDUH continues the 50 percent overlap by retaining half of the second-stage units from the 2011 survey. Those segments not retained are considered "retired" from use.

The first stage of the sample selection procedures began by geographically partitioning each State into roughly equal-sized State sampling (SS) regions. These regions were formed as a means of stratification so that each area would yield roughly the same expected number of interviews during each data collection period. This partitioning divided the United States into 900 SS regions made up of counties or groups and parts of counties.

The first stage of selection for the 2005–2012 surveys was census tracts. This stage of selection was included to contain sample segments within a single census tract to the extent possible.[1] Within each SS region, a sample of 48 census tracts was selected with probabilities proportional to size and with minimum replacement.

Because census tracts generally exceeded the minimum dwelling unit (DU) requirement,[2] selected census tracts were subdivided into smaller geographic areas of adjacent census blocks—called segments—that served as the second-stage sampling units. One segment per selected census tract or a total of 48 segments per SS region were selected (with probabilities proportional to size): 24 to field the 5-year study and 24 to serve as backups in case of sample depletion or to field any supplemental studies that the Substance Abuse and Mental Health Services Administration (SAMHSA) may request. For the 2012 survey, a total of 7,200 segments within the 900 SS regions were selected. Of the total, 3,600 segments were overlap segments used during the 2011 survey, 3,583 were new, and 17 were duplicates of segments used in the 2005–2012 surveys. For this last category, the same area had been listed previously or was being listed for the 2012 survey under a different segment identification number, so the original listing was used instead of relisting the same area.

After selecting these new areas, the process of counting and listing (C/L) the DUs within each new segment ensued. Segments to be used in 2012 were listed between April and December

[1] Some census tracts had to be aggregated to meet the minimum DU requirement.
[2] The minimum DU requirement was 150 DUs in urban areas and 100 DUs in rural areas.

2011. Once all DUs for a particular quarter were listed, the third-stage selection process identified sample dwelling units (SDUs) for inclusion in the study.

At the final stages of selection, five age-group strata were sampled at different rates. These five strata were defined by the following age-group classifications: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. No race/ethnicity groups were purposely oversampled for the 2012 survey. However, consistent with previous NSDUHs, the 2012 NSDUH was designed to oversample younger age groups by requiring equal sample sizes for the three age groups: 12 to 17, 18 to 25, and 26 or older.

2.2 Recruiting and Training for Counting and Listing

Preparations for C/L activities began with the decision to use the existing NSDUH data collection management structure to supervise counting and listing. All current field supervisors (FSs) were asked to handle the administrative tasks for the listers hired for their area. These tasks included completing the initial recruiting and hiring process, managing new lister mentoring and segment assignments, overseeing the timely completion of segments, and approving weekly time and expense reports. For technical supervision such as how to handle a specific segment, all listers contacted either the C/L manager or the Sampling Department at RTI International for answers and advice.

Beginning in March 2011, FSs recruited listing personnel from their existing staff of field interviewers (FIs). Experienced and new listers not currently working as FIs were also available for hire. A total of 383 were hired, certified, and worked from April through December 2011 to complete C/L operations for the 2012 NSDUH. Of the 383 listers, 291 worked as FIs on the 2011 NSDUH. In addition, 283 were returning listers from the 2011 C/L effort.

All hired listers received a certification training package containing a memorandum and materials including a C/L manual; Production, Time, and Expense report; hire letter; and instructions on accessing and completing four iLearning courses and a home study via the Internet. The four iLearning courses completed by all hired listers contained a lesson and assessment portion. The courses provided detailed training in topical areas such as listing multi-unit structures and group quarters, creating correct paths of travel, working efficiently, and avoiding common listing errors. Although the assessment portion was not graded, listers had to complete all four iLearning courses before completing the electronic home study. The home study included questions about C/L procedures as well as path-of-travel exercises. Both the iLearning courses and home study could be completed from any computer with Internet access. Hired listers who were not already working as FIs on NSDUH received an additional memorandum containing instructions on (1) completing a fifth iLearning course via the Internet that detailed the requirements of the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA) and (2) signing a Data Collection Agreement.

Listers had 2 weeks upon receipt of the certification training package to complete the certification process, which included reviewing the C/L manual; completing the four iLearning courses; passing the electronic home study with a score of 70 percent or higher on each of two sections; and completing the "CIPSEA Training" iLearning course and returning a signed Data Collection Agreement (for staff hired as listers only).

To work as a lister on NSDUH, all the requirements of the certification process had to be met. Of the 410 certification training packages distributed, five listers failed to complete the certification process within the 2-week time allotment. Another nine listers did not pass both sections of the electronic home study on their first attempt. Eight listers received feedback and retraining on questions missed and were given a second opportunity to retake the home study they failed. All eight listers passed on their second attempt. One newly hired lister was not given a second opportunity at the request of the FS and regional supervisor (RS) because of poor performance. In addition, 21 certified listers did not actually complete any listing work because of resignations or terminations.

All certified listers received their bulk listing supplies prior to beginning work. FSs assigned segments to certified listers via the web-based Case Management System (CMS), carefully considering the location and availability of their staff. After receiving their assigned segment materials packets, certified listers were then authorized to begin their C/L assignments. All listers sent their completed assignments directly to the Sampling Department, where the assignments were carefully edited. To improve the quality of the listing process, suggestions for improvement were provided to listers when necessary. Segments with significant errors were either refielded (for correction of major errors) or were corrected by sampling staff through discussions with the lister. In some cases, the lister returned to the segment to review the items in question.

2.3 Counting and Listing Procedures

Prior to the start of actual C/L field work, segment materials packets were assembled at RTI. Each packet contained maps of the selected area, listing forms, and blank segment information sheets. A copy of the maps remained at RTI for reference when assisting with problems encountered in the field.

Beginning in April, segment materials packets were assigned and sent to those listers who had completed the certification process and were ready to begin listing. Once the remaining staff became certified, they received assignments as well. Listers recorded the address or description of up to 400 DUs in each segment.

To reduce the time required to count and list segments, several procedures were implemented to maximize efficiency. In many cases the "count" step was eliminated: The lister could immediately list the segment unless, during the initial trip around the boundaries of the segment, it was apparent the segment had experienced additional construction or the lister determined that the segment was large (400+ DUs). As had been done on prior rounds of NSDUH, a rough count procedure was allowed for segments containing large geographic land areas, large DU counts (400+ DUs), or significant growth in residential DUs (typically, 1,000 DUs). This procedure permitted listers to obtain an approximate count of residential DUs in these segments from secondary sources—such as the post office, fire department, or county or city planning office—without having to conduct an exact count.

If a lister came across a segment that needed subsegmenting, the lister called in the initial DU counts to the Sampling Department, which could sometimes subsegment it over the telephone (any segment with 400+ DUs generally required subsegmenting). In cases involving

traveling listers, the telephone subsegmenting process allowed the lister to count and list a segment with 400 or more DUs in one trip. This was beneficial because a second trip would likely result in a delay of 1 or 2 weeks. For difficult subsegmenting tasks, the segment materials were sent to RTI to be handled directly by sampling personnel. Of the 3,583 new segments listed for the 2012 survey, 328 required subsegmenting. When obvious and possible, sampling staff completed any needed subsegmenting prior to the assignment of the segment to the lister, although the majority of subsegmenting occurred during the listing process. In a few cases where the subsegmenting was conducted prior to assignment of the segment to the lister, the census counts were outdated and the selected area was still too large to list. As a result, these areas had to be subsegmented a second time using field counts provided by the lister.

The counting and listing of all of the segments was completed by the end of December 2011. Once the segments were listed and the completed segment materials packets were received at RTI, an editing process of the completed materials checked for and deleted any DUs located outside segment boundaries, ensured that listing sheets matched segment sketches and maps, and verified that proper listing order and related listing rules were observed. During this editing process, the sampling staff also checked all subsegmenting that occurred in the field to ensure it was done correctly.

Listed DUs were keyed into a computer control system. A selection algorithm selected the specific SDUs to be contacted for the study. Prior to the beginning of the appropriate quarter, FSs assigned segments (or partial segments) to their interviewing staff. FIs received all assigned SDUs on their iPAQ handheld computer. Each selected unit and the next listed unit (for use as a sample check to capture missed DUs during screening and interviewing) were also printed on Selected DU Lists. These lists, along with copies of the handwritten listing forms and maps, were distributed to the assigned FI before the start of each quarter.

2.4 Added Dwelling Units

During the screening process, FIs were trained to identify any unlisted DUs that existed within the SDU or within the interval between the SDU and the next listed DU. If the missed DUs were housing units, they were automatically entered into the iPAQ (up to established limits) and selected for participation. At most, the FI could independently add 5 missed DUs per SDU and a maximum of 10 missed DUs per segment. If the FI discovered more than these amounts or if the missed DUs were group quarters units, the FI called their FS. The FS then either called RTI's Sampling Department for further instructions or instructed the FI to call the Sampling Department directly, depending on the situation.

Although no upper limit was placed on the total number of DUs that could be added to a segment by RTI's Sampling Department, the FIs were instructed to notify RTI of any significant listing problems. In a small number of segments, portions of these segments had to be relisted during the screening and interviewing phase. Table 2.2 indicates the number of segments that experienced added DUs, as well as the total number of added DUs for the 2012 NSDUH.

2.5 Problems Encountered

2.5.1 Controlled Access

In many of the major urban areas, listers had some difficulties gaining access to locked buildings and, in particular, had some trouble listing very large public housing complexes. Access in some suburban areas proved problematic as well; more and more planned communities have intercoms, guarded gatehouses, or entryways outfitted with cameras and scrambled buzzer systems. Access to military bases, college dormitories, boarding schools, and large retirement communities also proved problematic at times. Based on experience, these types of access problems were expected. Special mechanisms or protocols were in place to handle them promptly and, in some cases, avoid them entirely.

Access problems were typically resolved through effective follow-up efforts of supervisory staff, including situation-specific letters of request and in-person visits by the FSs and/or RSs. In particularly difficult situations, SAMHSA offered additional support via special refusal conversion letters or telephone follow-ups by the Project Officer.

2.5.1.1 Military Bases

As in past years, the often problematic access to military bases was handled with a formal and standardized approach for 2012. Through joint RTI and SAMHSA efforts, a contact person within the Pentagon for each branch of the service was identified. These individuals were advised in advance of base selections for the year. They then notified the base commanders regarding RTI's need to access these bases for both listing and screening and interviewing work. Additionally, RTI staff sent standard letters and informational packages to help obtain access to all selected bases. These efforts were effective: Access to the majority of the selected bases was secured.

2.5.1.2 Colleges and Universities

Access to colleges and universities is sometimes problematic. RTI used several standard approaches to accommodate the concerns of school administrators. Having standardized letters available that addressed recurring issues with a variety of attachment options was very effective.

Most schools requested or required only a letter stating the sponsor and the purpose of the study and identifying the lister or data collection staff. However, some schools wanted more complete information and the right to approve the field data collection procedures and personnel working in and around their campuses. Most of these situations resulted in packages being sent that contained the following:

1. RTI Institutional Review Board information;
2. Office of Management and Budget approval information;
3. descriptive information about the procedures and data collection plan; and
4. various study materials used with respondents during data collection.

In the end, the majority of private educational institutions expressing concerns cooperated in the C/L phase of the 2012 NSDUH.

2.5.2 Segments with Reassigned Quarters

Twenty-eight segments were identified during the C/L phase as difficult to access during months with unusual weather. Most access problems were due to roads being impassable because of snow during the winter months. Others involved roads being inaccessible because of rain, and one or two isolated locations involved water-only access that often froze during the winter months. If segments with weather or geographic access problems were selected for a quarter in which the access would be a problem (generally Quarters 1 or 4), the segment was switched with a segment in the same region for an appropriately paired time period. For example, inaccessible Quarter 1 segments were switched with Quarter 2 segments in the same region that would be more accessible during Quarter 1; Quarter 4 segments were switched with more easily accessed Quarter 3 segments. Generally the "switched" segment was selected because it had more accessible road surfaces, was more urban, or had fewer inaccessible roads.

In a few locations, such as some areas in Alaska, there were no segments that were better for reassignment during the problematic time period. When that happened, staff made prompt assignments, emphasized early completion of the work, and tried to plan around good weather forecasts to accomplish the field work as early in the period as possible.

2.5.3 Edited Addresses

In 2012, FIs continued to follow the detailed Editing Address Protocol initially implemented in Quarter 1 of 2006. This protocol emphasized the importance of exercising care when editing addresses, which in turn could alter the sample frame, particularly if the edit created a duplicate address.

FIs encountering a potential address change referred to a chart that listed various editing address scenarios, along with instructions to follow in each scenario.

Project management closely monitored reports on the web-based CMS for any potential problems resulting from address changes. A Duplicate Address report, updated daily, captured edited addresses made by FIs that produced duplicate listings. A separate Edited Address report, also updated daily, listed changes made to addresses other than those appearing on the Duplicate Address report.

As a result of the continued monitoring of edited addresses using the Editing Address Protocol, the incidence of problems potentially affecting the sampling frame was minimal. Any such problems were handled carefully by sampling staff to maintain the integrity of the NSDUH sample.

Table 2.1 2012 NSDUH Sampling Summary

Statistic	Small States	Big States	Total
Total Sample			
State Sampling Regions	516	384	900
Segments	4,128	3,072	7,200
Selected Lines	121,967	92,307	214,274
Eligible Dwelling Units	100,470	78,116	178,586
Completed Screening Interviews	89,618	64,255	153,873
Selected Persons	49,410	38,246	87,656
Completed Interviews	39,258	29,051	68,309
Average per State			
State Sampling Regions	12	48	
Segments	96	384	
Selected Lines	2,836	11,538	
Completed Interviews	913	3,631	
Interviews per Segment	9.51	9.46	
Average per State and Quarter			
Segments per State Sampling Region	2	2	
Interviews per State Sampling Region	19.02	18.91	
Interviews per Segment	9.51	9.46	
Total States	43	8	51
Total Interviewers (approximate number that varied by quarter)	501	335	836

Note: "Small" States refers to States where the design yielded **913** respondents on average. "Big" States refers to States where the design yielded **3,631** respondents on average.

Table 2.2 2012 Segments with Added Dwelling Units

Number of Added DUs per Segment (*X*)	Number of Segments with *X*-Added DUs	Cumulative Number of Added DUs*
1	585	585
2	152	889
3	71	1,102
4	31	1,226
5	16	1,306
6	11	1,372
7	3	1,393
8	7	1,449
9	3	1,476
10	3	1,506
11	2	1,528

*Total number of added dwelling units (DUs) = 1,528.

Exhibit 2.1 2012 NSDUH Sample Design Summary

First Stage of Selection for the NSDUH: Census Tracts

The 2005–2012 NSDUH design provided for estimates by State in all 50 States and the District of Columbia. States should therefore be viewed as the "first level" of stratification as well as a reporting variable. Eight States, labeled the "big" States in Table 2.1, had samples designed to yield 3,600 respondents per State. The remaining 43 "small" States[1] had samples designed to yield 900 respondents per State.

The larger sample sizes obtained at the State level, along with small area estimation techniques refined under previous NSDUH contracts, enabled the development of estimates for all States, for several demographic subgroups within each State (i.e., age group and race/ethnicity group), and for some core-based statistical areas (CBSAs) and a few small areas in the "big" States.

The "second level" of stratification defined contiguous geographic areas within each State. These State sampling (SS) regions were of approximately equal population size in terms of allocated samples.

Additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES (core-based statistical area/socioeconomic status) indicator[2] and by percentage of non-Hispanic white. The first-stage sample units for the 2005–2012 NSDUHs were selected from this well-ordered sample frame. Forty-eight census tracts per SS region were selected with probabilities proportionate to a composite size measure and with minimum replacement.

Second Stage of Selection for the NSDUH: Segments

For the second stage of sampling for the 2005–2012 NSDUHs, each of the selected census tracts was partitioned into noncompact clusters of dwelling units by aggregating adjacent census blocks. Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks were referred to as *segments*. On average, segments were formed so that they contained at least 150 dwelling units in urban areas and 100 dwelling units in rural areas and were constructed using 2000 Decennial Census data supplemented with revised population counts obtained from outside sources. A sample *dwelling unit* in NSDUH refers to either a housing unit or a group quarters unit (such as a dormitory room or a shelter bed).

One segment was selected within each selected census tract, with probability proportionate to size. Segments were formed so that they contained sufficient numbers of dwelling units to support three annual NSDUH samples. This allowed half of the segments used in any given year's sample to be used again in the following year as a means of improving the precision of measures of annual change. This also allowed for any special supplemental sample or field test that SAMHSA wished to conduct in any given NSDUH year within the same segments.

In order to coordinate the sample selection for 2005 through 2012, 48 census tracts were selected within each SS region, and 1 segment was selected per sampled census tract, for a total of 48 segments. An equal probability subsample of eight segments was used for the 2012 NSDUH. These eight segments were randomly assigned to quarters and to two panels within each quarter. The panels used in the 2012 NSDUH were designated as Panels 8 and 9. Panel 8 segments were used for the 2011 and 2012 surveys. New dwelling units (i.e., those not previously selected for the 2011 study) were selected from the Panel 8 segments for 2012. Panel 9 segments were new for 2012 and will be used again for the 2013 survey.

Exhibit 2.1 2012 NSDUH Sample Design Summary (continued)

Data from roughly one fourth of the final sample of respondents was collected during each calendar quarter. This important design feature helped control any seasonal bias that might otherwise exist in drug use prevalence estimates and other important NSDUH outcome measures of interest.

Third Stage of Selection for the NSDUH: Listed Lines

Before any sample selection within selected segments began, specially trained staff listed all dwelling units and potential dwelling units within each newly selected area segment. A dwelling unit is either a housing unit for a single household or one of the eligible noninstitutional group quarters that are part of the defined target population. The listings were based primarily on observation of the area segment and could include vacant dwelling units and units that appeared to be dwelling units but were actually used for nonresidential purposes. The objective of the listing was to attain as complete a listing as possible of eligible residential addresses; any false positives for residences were eliminated during the household screening process after the sample was selected.

The sampling frame for the third stage of sample selection was the lines of listed dwelling units and potential dwelling units. After accounting for eligibility, nonresponse, and the fourth-stage sample selection procedures, it was determined that 198,370[3] lines were needed to obtain a sample of 67,500 responding persons distributed by State and age group. During the study's implementation, however, a total of 214,274 lines were selected and yielded a final respondent sample of 68,309 (as shown in Table 2.1).

As in previous years, if a field interviewer encountered any new dwelling unit in a segment or found a dwelling unit missed during the counting and listing activities, the new and missed dwelling units were selected into NSDUH using a half-open interval selection technique.[4] That selection technique eliminated any frame bias that might have been introduced because of errors and/or omissions in counting and listing activities and also eliminated any bias that might have been associated with using "old" segment listings.

Fourth Stage of Selection for the NSDUH: Persons

After dwelling units were selected within each segment, a field interviewer visited each selected dwelling unit to obtain a roster of all persons aged 12 or older residing in the dwelling unit. This roster information was then used to select zero, one, or two persons for the survey. Sampling rates were preset by age group and State. Roster information was entered directly into the electronic screening instrument (the iPAQ), which automatically implemented this fourth stage of selection based on the State and age group sampling parameters.

Using an electronic screening instrument also provided the ability to impose a more complicated person-level selection algorithm at the fourth stage of selection. As a result of this unique design feature, *any* two survey-eligible persons within a dwelling unit had some chance of being selected—that is, all survey-eligible pairs of persons had some non-zero chance of being selected. This design feature is of interest to NSDUH researchers because it allows analysts to examine how the drug use propensity of one individual in a family relates to that of other family members residing in the same dwelling unit (e.g., the relationship of drug use between a parent and child). Originally added in 2002 with use continuing through 2012, an additional parameter in the person selection process increased the number of selected pairs within dwelling units without unduly diminishing response rates.

As illustrated in Table 2.1, at the fourth stage of selection, 87,656 persons were selected from 153,873 screened and eligible dwelling units. A total of 68,309 completed interviews were obtained from these 87,656 selected persons.

Exhibit 2.1 2012 NSDUH Sample Design Summary (continued)

Expected Precision of NSDUH Estimates

The multistage, stratified NSDUH design has been optimally constructed to achieve specified precision for various person subpopulations of interest. These SAMHSA-specified precision requirements call for the expected relative standard error on a prevalence of 10 percent not to exceed the amounts listed below.

For the **NSDUH**:

- 3.00 percent for total population statistics; and
- 5.00 percent for statistics in three age group domains: 12-17, 18-25, and 26 or older.

To achieve these precision requirements and meet State sample-size requirements, the optimal person-level sample distribution by strata was determined. This sample distribution minimized data collection costs while simultaneously meeting the above-specified precision requirements for several critical NSDUH outcome measures.

[1] For reporting and stratification purposes, the District of Columbia is treated the same as a State and no distinction is made in the discussion.

[2] The four categories are defined as (1) CBSA/low SES, (2) CBSA/high SES, (3) Non-CBSA/low SES, and (4) Non-CBSA/high SES.

[3] See the 2012 sample design plan (Morton, Hirsch, & Martin, 2011).

[4] In summary, this technique states that if a dwelling unit is selected for NSDUH and a field interviewer observes any new or missed dwelling units between the selected dwelling unit and the dwelling unit appearing immediately after the selection on the counting and listing map page, then all new and missed dwellings between the selection and the next one listed will be selected. If a large number of new and missed dwelling units are encountered (generally greater than 10), then a sample of the missed dwelling units will be selected.

References

Morton, K. B., Hirsch, E. L., & Martin, P. C. (2011). *2012 National Survey on Drug Use and Health: Sample design plan* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase II, Deliverable No. 7, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

3. Data Collection Staffing

The magnitude of the National Survey on Drug Use and Health (NSDUH) required a field data collection management structure robust enough to support the interviewing staff and flexible enough to manage an ever-changing variety of issues. The basic management structure remained unchanged from prior surveys: field supervisors (FSs) managed States and substate regions and reported to regional supervisors (RSs) who then reported to regional directors (RDs) who reported directly to the National Field Director. This chapter discusses the process of staffing the 2012 NSDUH data collection effort.

3.1 Regional Directors

The RDs managed data collection within defined territories of the Nation. Reporting directly to the National Field Director, the RDs, working with the project director and the National Field Director, served as the management team for all data collection operations.

In 2012, the Nation was divided among three RDs for data collection. The RDs in place at the end of 2011 continued their roles on the 2012 NSDUH. All RDs were survey managers with many years of experience at RTI and on NSDUH.

Each of the RDs managed a staff of RSs, who in turn managed a staff of five or six FSs who managed the team of field interviewers (FIs) in their individual States or assigned areas. Each RD worked with the traveling field interviewer (TFI) manager who coordinated the work of TFIs within the RD's region.

RDs also had project-wide ancillary functions not specific to their region. These included coordinating controlled access communications, FS recruiting, and TFI manager work.

Exhibit 3.1, at the end of this chapter, displays the RD regions and management task assignments at the end of the 2012 NSDUH. Listed under each RD is the structure containing the number of RSs and FSs, geographic regions, and the ancillary management functions.

3.2 Regional Supervisors

The RSs were the direct managers of five or six FSs. Reporting to an RD, RSs were responsible for all data collection activities in the States in their region. The States, including the District of Columbia, were clustered geographically to be managed by the RSs. Of the seven RSs on the supervisory team at the start of 2012, six had served as RSs throughout the 2011 survey, and one began in Quarter 4, 2011. In Quarter 2, 2012, one RS transferred to another RTI project. This vacancy was filled by an RTI survey specialist with previous experience on NSDUH. See Exhibit 3.1 for the final groupings of States managed by each RS.

3.3 Field Supervisors

The FSs were the first-level supervisors of FIs conducting data collection in each of the States. The FSs assigned work, monitored progress, resolved problems, and managed the day-to-day activities of their FIs. Each FS reported directly to an RS.

In addition, a substitute FS was available to provide coverage for FSs who were on vacation or experiencing emergencies. The substitute FS also helped with FI recruiting, problem resolution, and mentoring of new FIs as needed. If multiple regions requested assistance at the same time, project management assessed where the greatest needs were and assigned the substitute FS accordingly.

At the end of 2012, there were 40 FSs (see Exhibit 3.1).

3.4 Field Interviewers and Traveling Field Interviewers

One of the primary FS functions was the continuous recruiting and hiring of the FI staff needed to complete the data collection work each quarter. Subcontractor Headway Corporate Resources is the staffing agency serving as the employer of record for all FIs hired for the NSDUH. FSs worked with Headway's Center for Operational & Recruitment Excellence (CORE) to identify potential FI candidates. Multiple recruiting approaches were used, including:

- reviewing Headway's Interviewer Database, which contains information of previous RTI interviewers who are eligible for rehire, as well as candidates from previous recruiting efforts who were considered qualified but not hired;
- networking;
- placing newspaper advertisements and posting informative job flyers;
- providing recruiting business cards;

- contacting job service agencies; and
- using Internet job advertising and search services.

Networking involved any or all of the following contacts:

- other FSs;
- RTI staff working on other surveys with potential FIs available;
- other survey research organizations; and
- other FIs (current NSDUH FIs recommending successful candidates received a recruiting bonus).

Those with general interviewing experience, and especially those with experience working on government surveys, were given preference in hiring. However, candidates with transferable skills and experience—such as contact with the public, attention to detail, and organizational skills—were considered.

The work of an interviewer requires a wide range of skills and abilities. Some of the characteristics and qualities that FSs tried to identify in potential hires included:

- intelligence;
- dependability;
- sensitivity and objectivity;
- honesty;
- ability to follow instructions;
- reading ability;
- listening skills;
- motivation;
- availability; and
- flexibility.

In order to make an informed decision, potential hires also needed to find out more about the role of an FI on NSDUH. Comprehensive and realistic information packets, which included a video and other materials about being an FI, were emailed to interested persons.

FSs then contacted applicants over the telephone to determine if an in-person interview was warranted. Viable FI candidates still interested in the job were interviewed by the FS using behavior-based questions that required the candidates to provide examples about how they had handled specific situations in the past. For example, an FS might say, "Tell me about the last time you were in a situation where you had to approach a stranger to extract some sort of information. How did you do it?" Also during the interview, the FS fully explained the requirements and responsibilities of the NSDUH FI's job, described the project expectations, and defined the required time commitment. The FS then probed the candidate's job and interviewing

history. The FS completed reference checks for candidates scheduled for an in-person interview. At the conclusion of the interview, if the FS still considered the person a viable FI candidate and the reference checks were satisfactory, the FS then recommended the candidate for hire. Criminal background and driving history checks were then completed before the candidate attended a training session.

At each new-to-project (NTP) interviewer training session during 2012, fingerprint impressions were collected from all newly hired FIs for further investigation by the U.S. Office of Personnel Management (OPM) and the Federal Bureau of Investigation (FBI). This was a requirement for employment, and any FIs who chose not to have fingerprints taken were ineligible for employment as a NSDUH FI.

It was essential that staff hired to serve as FIs understood and were committed to the standards of confidentiality and excellence required by NSDUH. To help ensure this, all individuals hired to serve as FIs were required to read and sign a Data Collection Agreement (see Exhibit 3.2). Failure to comply with the provisions of this agreement would have resulted in termination from NSDUH.

FSs attempted to hire bilingual interviewers who spoke Spanish fluently in those sample areas with large Spanish-speaking populations. Before an FS hired a bilingual candidate, each applicant was screened by a bilingual staff member to assess the applicant's English- and Spanish-language abilities. The assessment involved reading and speaking in English and Spanish. The bilingual candidate had to meet these assessment requirements satisfactorily before he or she could be hired and trained as an RTI-certified bilingual interviewer.

Another subset of specialized interviewers was the TFIs. Each RD region had access to a team of TFIs with proven interviewing experience.

Each TFI was asked to commit to at least two 12-day trips each quarter. TFI teams were used to fill the unmet needs in areas with staffing shortfalls or where special needs arose (such as covering long-term illnesses among the staff). In addition, two TFIs were certified bilingual interviewers and were assigned to areas where no bilingual interviewer was available. During 2012, the TFI team consisted of 11 active interviewers.

Exhibit 3.3 displays a flow chart that presents all of the steps in the FI recruiting and hiring process.

During the entire data collection period, a total of 836 FIs completed training and worked on the study. The following are demographic characteristics of the interviewing staff:

- Of the total 836 FIs, 640 (76.5 percent) were veteran interviewers who had worked on the 2011 NSDUH, while 196 (23.5 percent) were newly hired and trained during 2012.
- Of the total 836 FIs, 590 (70.6 percent) were white; 82 (9.8 percent) were black or African American; 164 (19.6 percent) identified themselves as "Other" (including Asian, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, etc.); and 118 (14.1 percent) were bilingual in Spanish.

At the end of this chapter, Table 3.1 provides a distribution of interviewers by race and gender for the veteran interviewers, Table 3.2 for the interviewers hired and trained during 2012, and Table 3.3 for the total. Table 3.4 provides a distribution of veteran interviewers by bilingual skill and gender, Table 3.5 for the newly trained interviewers, and Table 3.6 for the total.

3.5 Problems Encountered

3.5.1 Continued Staffing Shortfall in Certain Areas

In certain areas, the number of staff working continued to be less than the targeted number of FIs needed. This targeted number was based on:

- allocation of the sample across the FI regions each quarter;
- number of hours that an average FI would work each week, based on recent experience;
- average length of time to complete each screening;
- average length of time to complete each interview; and
- number of weeks that the interviewing staff would work in the quarter based on recent experience.

As each quarter's sample was provided by the statisticians, the process to estimate the number of needed FIs was repeated. The assumptions were refined based on the most recent experience. The number of FIs needed from quarter to quarter varied, so FSs had to review staff assignments throughout the quarter and continually recruit and hire additional FIs.

While most areas were close to the targeted number, some areas struggled. To compensate for these problem areas, TFIs were used to perform the work. Supervisors also borrowed FIs from other areas to complete the work. These borrowed FIs had completed their initial assignment and were willing to travel and take on additional work.

3.5.2 FI Turnover

In 2012, the turnover[1] rate among FIs was 19.4 percent, a decrease from 24 percent in 2011. The continuing FI turnover meant Headway's CORE group and FSs had to continually recruit new staff, and FSs had to juggle assignments to ensure that all of the work was completed appropriately. There were significant costs associated with continuous recruiting efforts. These included not only the time of the FSs and the RTI office staff, but traveling to conduct interviews with candidates, and eventually training the newly hired staff. Additional costs were also incurred when TFIs had to be sent to work in areas where no FI was available.

To combat FI turnover, RTI and Headway's CORE group took a variety of steps, including:

[1] FI turnover rate was referred to as "attrition rate" in reports prior to 2008. The calculations for this rate remain unchanged; the terminology has been changed to more accurately describe these calculations.

- recruiting and carefully selecting qualified staff who understood the demands of the job before being hired;
- training staff thoroughly and mentoring all new staff in the field;
- supporting staff with individual calls at least once each week and group calls at least once each quarter; and
- providing assurance of never being alone: there is always someone to call for assistance.

Table 3.1 Distribution of 2012 Veteran Interviewers, by Race and Gender

	Male		Female		Total	
Race	**Count**	**%**	**Count**	**%**	**Count**	**%**
Black or African American	9	6.4	52	10.4	61	9.5
White	102	72.9	364	72.8	466	72.8
Other	29	20.7	84	16.8	113	17.7
Total	140	100.0	500	100.0	640	100.0

Table 3.2 Distribution of Interviewers Hired in 2012, by Race and Gender

	Male		Female		Total	
Race	**Count**	**%**	**Count**	**%**	**Count**	**%**
Black or African American	1	2.6	20	12.7	21	10.7
White	24	61.5	100	63.7	124	63.3
Other	14	35.9	37	23.6	51	26.0
Total	50	100.0	157	100.0	196	100.0

Table 3.3 Distribution of All 2012 Interviewers, by Race and Gender

	Male		Female		Total	
Race	**Count**	**%**	**Count**	**%**	**Count**	**%**
Black or African American	10	5.6	72	11.0	82	9.8
White	126	70.4	464	70.6	590	70.6
Other	43	24.0	121	18.4	164	19.6
Total	192	100.0	657	100.0	836	100.0

Table 3.4 Distribution of 2012 Veteran Interviewers, by Gender and Language Ability

	Male		Female		Total	
Language Ability	**Count**	**%**	**Count**	**%**	**Count**	**%**
Bilingual	23	16.4	58	11.6	81	12.7
Nonbilingual	117	83.6	442	88.4	559	87.3
Total	140	100.0	500	100.0	640	100.0

Table 3.5 Distribution of Interviewers Hired in 2012, by Gender and Language Ability

	Male		Female		Total	
Language Ability	**Count**	**%**	**Count**	**%**	**Count**	**%**
Bilingual	13	33.3	24	15.3	37	18.9
Nonbilingual	26	66.7	133	84.7	159	81.1
Total	39	100.0	157	100.0	196	100.0

Table 3.6 Distribution of All 2012 Interviewers, by Gender and Language Ability

	Male		Female		Total	
Language Ability	**Count**	**%**	**Count**	**%**	**Count**	**%**
Bilingual	36	20.1	82	12.5	118	14.1
Nonbilingual	143	79.9	575	87.5	718	85.9
Total	179	100.0	657	100.0	836	100.0

Exhibit 3.1 NSDUH Management Chart

Person names have been removed from chart.

Project 0212800

2012 NSDUH

NSDUH Project Organization

Quarter 4

Task Definitions

x01.001	Project Management	x07.006	Imputation Activities
x01.002	Consultant Meetings	x08.001	Detailed Tables
x02.001	Instrumentation Development	x08.002	6-Month Detailed Tables
x02.002	Training & Field Materials	x08.003	Editing & Imputation Eval
x02.003	Field Test	x08.004	National Findings Report
x03	Sample Design	x08.005	Sampling Error Report
x04	Sample Selection	x08.006	Ad Hoc Methods Analyses
x04.001	Mapping & Segment Kits	x08.007	Ad Hoc Rapid Analyses
x04.002	GIS Support for Mapping	x08.008	Ad Hoc Analytic Reports
x04.003	Edit & Prep for Interview	x09.001	State Analytic Report
x04.004	Field Support & Master Filing	x09.002	Substate Analytic Report
x05.001	Field Preparations	x10	Data Files and Documentation
x05.002	Field Training	x11	Survey Documentation
x06.001	Data Quality and Verification	x12.001	MHSS: Project Management
x06.002	Screen and Interview	x12.002	MHSS: Instrument Assessment
x06.003	Field Observations	x12.003	MHSS: Sample Design
x07.001	Data Management	x12.005	MHSS: Training & Field Prep
x07.002	Tech Support & Equipment Maint	x12.006	MHSS: Data Collection
x07.003	Data Editing & Other Spec	x12.007	MHSS: Data Management
x07.004	Weights and Adjustments	x12.008	MHSS: Analysis & Reporting
x07.005	Pair Weights and Adjustments	x12.010	MHSS: Files & Documentation

Exhibit 3.1 NSDUH Management Chart (continued)

Task Definitions

x01.001	Project Management	x07.006	Imputation Activities
x01.002	Consultant Meetings	x08.001	Detailed Tables
x02.001	Instrumentation Development	x08.002	6-Month Detailed Tables
x02.002	Training & Field Materials	x08.003	Editing & Imputation Eval
x02.003	Field Test	x08.004	National Findings Report
x03	Sample Design	x08.005	Sampling Error Report
x04	Sample Selection	x08.006	Ad Hoc Methods Analyses
x04.001	Mapping & Segment Kits	x08.007	Ad Hoc Rapid Analyses
x04.002	GIS Support for Mapping	x08.008	Ad Hoc Analytic Reports
x04.003	Edit & Prep for Interview	x09.001	State Analytic Report
x04.004	Field Support & Master Filing	x09.002	Substate Analytic Report
x05.001	Field Preparations	x10	Data Files and Documentation
x05.002	Field Training	x11	Survey Documentation
x06.001	Data Quality and Verification	x12.001	MHSS: Project Management
x06.002	Screen and Interview	x12.002	MHSS: Instrument Assessment
x06.003	Field Observations	x12.003	MHSS: Sample Design
x07.001	Data Management	x12.005	MHSS: Training & Field Prep
x07.002	Tech Support & Equipment Maint	x12.006	MHSS: Data Collection
x07.003	Data Editing & Other Spec	x12.007	MHSS: Data Management
x07.004	Weights and Adjustments	x12.008	MHSS: Analysis & Reporting
x07.005	Pair Weights and Adjustments	x12.010	MHSS: Files & Documentation

Exhibit 3.1 NSDUH Management Chart (continued)

Task Definitions

x01.001	Project Management	x07.006	Imputation Activities
x01.002	Consultant Meetings	x08.001	Detailed Tables
x02.001	Instrumentation Development	x08.002	6-Month Detailed Tables
x02.002	Training & Field Materials	x08.003	Editing & Imputation Eval
x02.003	Field Test	x08.004	National Findings Report
x03	Sample Design	x08.005	Sampling Error Report
x04	Sample Selection	x08.006	Ad Hoc Methods Analyses
x04.001	Mapping & Segment Kits	x08.007	Ad Hoc Rapid Analyses
x04.002	GIS Support for Mapping	x08.008	Ad Hoc Analytic Reports
x04.003	Edit & Prep for Interview	x09.001	State Analytic Report
x04.004	Field Support & Master Filing	x09.002	Substate Analytic Report
x05.001	Field Preparations	x10	Data Files and Documentation
x05.002	Field Training	x11	Survey Documentation
x06.001	Data Quality and Verification	x12.001	MHSS: Project Management
x06.002	Screen and Interview	x12.002	MHSS: Instrument Assessment
x06.003	Field Observations	x12.003	MHSS: Sample Design
x07.001	Data Management	x12.005	MHSS: Training & Field Prep
x07.002	Tech Support & Equipment Maint	x12.006	MHSS: Data Collection
x07.003	Data Editing & Other Spec	x12.007	MHSS: Data Management
x07.004	Weights and Adjustments	x12.008	MHSS: Analysis & Reporting
x07.005	Pair Weights and Adjustments	x12.010	MHSS: Files & Documentation

Exhibit 3.1 NSDUH Management Chart (continued)

Task Definitions

x01.001	Project Management	x07.006	Imputation Activities
x01.002	Consultant Meetings	x08.001	Detailed Tables
x02.001	Instrumentation Development	x08.002	6-Month Detailed Tables
x02.002	Training & Field Materials	x08.003	Editing & Imputation Eval
x02.003	Field Test	x08.004	National Findings Report
x03	Sample Design	x08.005	Sampling Error Report
x04	Sample Selection	x08.006	Ad Hoc Methods Analyses
x04.001	Mapping & Segment Kits	x08.007	Ad Hoc Rapid Analyses
x04.002	GIS Support for Mapping	x08.008	Ad Hoc Analytic Reports
x04.003	Edit & Prep for Interview	x09.001	State Analytic Report
x04.004	Field Support & Master Filing	x09.002	Substate Analytic Report
x05.001	Field Preparations	x10	Data Files and Documentation
x05.002	Field Training	x11	Survey Documentation
x06.001	Data Quality and Verification	x12.001	MHSS: Project Management
x06.002	Screen and Interview	x12.002	MHSS: Instrument Assessment
x06.003	Field Observations	x12.003	MHSS: Sample Design
x07.001	Data Management	x12.005	MHSS: Training & Field Prep
x07.002	Tech Support & Equipment Maint	x12.006	MHSS: Data Collection
x07.003	Data Editing & Other Spec	x12.007	MHSS: Data Management
x07.004	Weights and Adjustments	x12.008	MHSS: Analysis & Reporting
x07.005	Pair Weights and Adjustments	x12.010	MHSS: Files & Documentation

Exhibit 3.1 NSDUH Management Chart (continued)

FS = Field Supervisor

Exhibit 3.1 NSDUH Management Chart (continued)

Exhibit 3.1 NSDUH Management Chart (continued)

(S) = secondary report

TFI = Traveling Field Interviewer

Exhibit 3.2 Data Collection Agreement

headway workforce solutions ***DATA COLLECTIONAGREEMENT***	Project Name: **National Survey on Drug Use and Health** Project No.: **0212800**

I, _____, an employee of Headway, agree to provide field data collection services for the benefit of RTI in connection with the RTI Project shown above ("the Project"). Further, I

1) am aware that the research being conducted by RTI is being performed under contractual arrangement with the **Substance Abuse and Mental Health Services Administration** (SAMHSA);

2) hereby accept all duties and responsibilities of performing specified data collection tasks and will do so **personally,** in accordance with the training and guidelines provided to me. At no time will I engage the services of another person to perform any data collection tasks for me without the prior written approval of both my employer (Headway) and RTI;

3) agree to treat as **confidential** all information secured during interviews or obtained in any Project-related way during the period I am working on the Project, as required by the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), and understand, under Section 513 of this Act, I am subject to criminal felony penalties of imprisonment for not more than five years, or fines of not more than $250,000, or both, for voluntary disclosure of confidential information. Any breach of confidentiality must be reported immediately to the National Field Director. This information will be shared with the SAMHSA Project Officer and Headway. I have also completed and fully understand the CIPSEA training provided to me;

4) agree to treat as **confidential and proprietary** to RTI/SAMHSA any and all information provided by the public, whether collected or accessed in electronic or printed form during the course of my service on this Project, including but not limited to all data collection computer software and respondent data, and will protect such items from unauthorized use or disclosure;

5) am aware that the survey instruments completed form the basis from which all analyses will be drawn, and therefore, agree that all work for which I submit invoices will be legitimate, of high quality and performed in compliance with all Project specifications to ensure the scientific integrity of the data;

6) understand that I am fully and legally responsible for taking all reasonable and appropriate steps to ensure that any computer equipment issued to me for use on this Project is safeguarded against damage, loss, or theft. I also understand that I have a legal obligation to immediately return all equipment at the conclusion of my assignment or at the request of my supervisor;

7) fully agree to conduct myself at all times in a manner that will obtain the respect and confidence of all individuals that I encounter as a representative of the Project and I will not betray this confidence by divulging information obtained to anyone other than authorized Project representatives of RTI;

8) understand that evidence of **falsification, fabrication or distortion of any data** collected for this Project will be reported to RTI's Scientific Integrity Committee, and such acts are grounds for immediately removing me from the Project and can result in my suspension from any government-funded research. Also, if falsification of data is substantiated, I understand a **formal fraud complaint** will be submitted to the U.S. Department of Health and Human Services' **Office of Inspector General** (OIG) and I could be subject to **criminal and/or civil prosecution** and thereby face imprisonment, financial penalties or both;

9) understand my obligations under this agreement supersede any prior or existing agreements on the same subject matter and will survive the termination of any assignment with RTI and/or my employment by Headway.

_____ _____

Employee Signature **Date**

Exhibit 3.3 Flow of FI Recruiting Activity

*Occasionally, the requested background check information is not returned to RTI/Headway by the time the hire letter must be sent. In these instances, the hire letter states that employment is contingent upon the successful completion of the background check. All background checks are completed before new hires attend training.

4. Preparation of Survey Materials

RTI and Substance Abuse and Mental Health Services Administration (SAMHSA) staff preparing survey materials for the 2012 National Survey on Drug Use and Health (NSDUH) reexamined and updated the computer-assisted interviewing (CAI) program, the iPAQ electronic screening program, as well as all other manuals and interview materials. With veteran and new field interviewer (FI) training sessions, the preparation for training required meticulous planning.

4.1 Electronic Screening

Using the 2011 electronic screening program, changes were made to prepare the 2012 iPAQ screening program. Exhibit 4.1, at the end of this chapter, contains a complete list of changes from 2011 for the 2012 electronic screening program.

4.2 Questionnaire Development

4.2.1 CAI Instrument

Using the 2011 computer program, a number of changes were made to prepare the 2012 CAI instrument. Exhibit 4.2 contains a detailed list of all changes implemented between the 2011 and 2012 instrument versions.

Corresponding audio WAV files were recorded for all new items within the audio computer-assisted self-interviewing (ACASI) portion of the interview. Materials used during the actual interview, including the Reference Date Calendar, the Pill Cards, and the Showcard Booklet, were also updated.

4.2.2 Spanish Translations

Using the 2011 Spanish CAI instrument, the changes in the questionnaire and interview materials referred to above were translated and incorporated. Additional Spanish audio WAV files were recorded as well to allow respondents to listen to the ACASI sections in Spanish when necessary.

4.3 Manuals and Miscellaneous Materials Development

4.3.1 Manuals

Based upon the 2011 manuals, updated versions of the manuals listed below were prepared. These new versions provided all staff, both experienced and new, with accurate, detailed manuals for both training and reference:

- Field Interviewer Manual: All FIs received an FI Manual detailing all aspects of an interviewer's work requirements on the 2012 NSDUH. This manual was sent to all FIs for review prior to attending a new-to-project (NTP) or veteran FI training. It was

utilized throughout the training sessions and served as a ready reference when questions arose during fieldwork throughout the year. FIs were also able to access an electronic version of the 2012 FI Manual directly from the CAI Manager on the laptop computer. For supervisory and management staff, the FI Manual was available for reference on the web-based Case Management System (CMS). Veteran FIs were also provided a reference sheet listing important changes made to the manual for 2012.

- Field Interviewer Computer Manual: This companion FI manual provided details about hardware use and care issues for both the iPAQ and the Gateway laptop computer, instructions for using the programs on each computer, transmission steps, and a troubleshooting guide to assist staff encountering technical difficulties. This computer manual was included with—but bound separately from—the FI Manual, so FIs could easily include it in their computer carrying case as a quick reference while working. In 2012, FIs received a copy of the computer manual along with the 2012 FI Manual. An electronic version of the computer manual was also available on the CMS for supervisory and management staff.
- Field Supervisor Manual: This detailed manual for field supervisors (FSs) included instructions and tips for recruiting field staff and managing the counting and listing (C/L) effort and screening and interviewing work. Strategies for managing staff using information on the CMS were also presented, as were administrative issues for both the FSs and their staff. FSs, regional supervisors (RSs), and regional directors (RDs) were able to reference this manual on the CMS.
- Field Supervisor Computer Manual: Explanations of the equipment provided for FSs (computer, all-in-one printer, and peripherals) were included in this separate volume along with instructions on using the various software tools (Windows/Microsoft Word/Microsoft Excel, e-mail, FedEx tracking, etc.). Detailed instructions on how to use the CMS were provided for instruction and reference. FSs, RSs, and RDs were able to reference this manual on the CMS.
- Regional Supervisor Manual: This manual provided specific guidelines for RSs on supervising the FSs in their region and reporting requirements to the RDs. Separate chapters provided instructions for managing the various stages of NSDUH, including FI recruitment, C/L, and screening and interviewing. RSs and RDs were able to reference this manual on the CMS.
- Counting and Listing Manual: This manual included explanations and examples of the detailed C/L procedures. All listers received copies of the manual. Supervisory and management staff working on the C/L phase of NSDUH were able to reference this manual on the CMS.
- Data Quality Coordinator and Consistency Check Manuals: These manuals documented the processes to be followed by the Data Quality Team in the verification process and in resolving consistency check problems.
- Guide to Controlled Access Situations: This manual, available to all management staff, documented the various ways to try to gain admittance in challenging access situations.

- NSDUH Best Practices Guidebook: This guidebook for project management staff provided details about issues such as chain of command, use of the project network drive, whom to include on various e-mails, and various other specific project-related procedures, protocols, and activities.

4.3.2 Miscellaneous Materials

Based on the 2011 versions, the following respondent materials were updated for 2012:

- Reference Date Calendar;
- Interview Payment Receipt; and
- Summary of Questionnaire.

Minor modifications from the 2011 versions were made to the following respondent materials:

- Lead Letter (updated the survey year; updated dates on the FI ID Badge image);
- Study Description (updated the survey year);
- Intro to CAI for 12-17 (clarified wording of FI instructions above the parental permission script at the top of the page to read, "FIRST, READ THE PARENTAL PERMISSION SCRIPT BELOW AND OBTAIN PERMISSION FROM THE PARENT.");
- Quality Control Form (updated the survey year; lightened the MM/DD text included in the interviewer portion to be less intrusive);
- RTI/SAMHSA Fact Sheet (updated the list of international RTI office locations; updated SAMHSA information and revised the Access to Recovery grant total);
- Question and Answer (Q&A) Brochure (updated the list of primary objectives for NSDUH to be consistent with other NSDUH materials; revised the text to state the interview takes "about an hour");
- NSDUH Respondent Website (updated links to information on NSDUH; updated the NSDUH in the News information Web page with more recent articles and reports and added a video containing footage from the most recent NSDUH press conference; updated the list of primary objectives for NSDUH to be consistent with other NSDUH materials);
- NSDUH Highlights (updated text to reflect the 2010 study results);
- Showcard Booklet (updated the survey year; added additional dosages of Morphine ER (extended release) and Morphine IR (immediate release) to Pill Card A; added an Informed Consent Reference Guide to the Job Aids section for easy reference; updated text and FI Manual references in the Steps to Maximizing Data Quality exhibit in the Job Aids section to reflect current information; added reminders about plugging in the headphones while setting up the computer and tasks to complete during the ACASI to the Interview Process exhibit in the Job Aids section);

- Appointment Card (simplified text; re-formatted layout to present a more professional appearance);
- "Sorry I Missed You" Card (re-formatted layout to present a more professional appearance);
- Doorperson Letter (added name and signature; updated dates on the FI ID Badge image; edited the letterhead to make it consistent with the Lead Letter);
- Interview Payment Receipt (updated the year; updated information on SAMHSA's Treatment Referral Helpline);
- Newspaper Article handout (updated with two new articles featuring information on Methamphetamine and marijuana use); and
- SAMHSA Authorization Letter (updated the survey year and contract number).

For 2012, one NSDUH short report, *The NSDUH Report: Illicit Drug Use Among Older Adults* (Center for Behavioral Health Statistics and Quality, 2011a), and two Data Spotlights: *OAS Data Spotlight: 12 Year Olds More Likely to Use Inhalants Than Cigarettes or Marijuana* (Office of Applied Studies, 2010) and *Data Spotlight: Young Alcohol Users Often Get Alcohol from Family or Home* (Center for Behavioral Health Statistics and Quality, 2011b), were available for distribution to interviewers.

The following respondent materials remained virtually unchanged from 2011 for use in 2012:

- Certificate of Participation;
- Controlled Access Letters;
- Intro to CAI for 18+;
- Other Language Introduction Card;
- Refusal and Unable to Contact Letters;
- Spanish Card; Verification Letter (English and Spanish versions); and
- Who Uses the Data handout.

4.4 Submission of the 2012 NSDUH IRB Package

Once the 2012 survey materials, CAI program, and iPAQ screening program were finalized, these items were submitted to RTI's Institutional Review Board (IRB) as part of the IRB package for the 2012 NSDUH. RTI's IRB Committee met on September 20, 2011.

During the IRB Committee review of the package, the Committee requested additional information about the process for translating the survey instruments and materials into Spanish. After submitting a response describing the process on September 22, 2011, IRB approval for the 2012 NSDUH was received on September 27, 2011.

In addition, a series of addendums was submitted to the IRB for the 2012 NSDUH in response to an issue that arose during the Committee review of the 2013 NSDUH package on September 26, 2012. During the review of the 2013 NSDUH, the Committee required revisions to the informed consent procedures for youths who are 17 years old and living independently without a parent or guardian residing in the home. As these same procedures were also used for the 2012 NSDUH, the Committee required similar changes for the 2012 survey.

In response, an addendum was submitted to the IRB on October 5, 2012 for the 2012 and 2013 NSDUHs to formally request a waiver of parental permission for 17-year-old respondents living independently without a parent or guardian. Along with the addendum, documentation was provided to reflect required changes to informed consent procedures and project materials, specifying that the only situation where parental permission was not required was for youths who are 17 years old and living independently without a parent or guardian residing in the home. For all youths aged 16 or younger, parental permission was required with no exceptions. Communication with field staff on this issue was also included with the addendum. Approval of the addendum was received on October 22, 2012.

Two addendums related to the Mental Health Surveillance Study (MHSS) were submitted to the IRB during 2012. The first addendum, dated February 27, 2012, added text to the Distressed Respondent Protocol (DRP) instructing the clinical interviewer to ask respondents if they had a doctor or other professional they could speak with and to contact emergency providers if warranted by the situation. The second addendum, dated November 15, 2012, requested an update to the SAMHSA website address provided in the DRP and used by clinical interviewers to locate the nearest emergency psychiatric service. Both addendums were approved.

4.5 Preparation for New-to-Project Field Interviewer Training

This section reviews the main steps necessary to prepare for NTP FI trainings.

4.5.1 Home Study Package

Prior to training, each new FI hired for screening and interviewing work was sent a home study package containing:

- 2012 Field Interviewer Manual;
- 2012 Field Interviewer Computer Manual;
- 2012 Mental Health Surveillance Study FI Handbook;
- cover memorandum from the National Field Director;
- paper version of the Electronic Home Study Exercises; and
- background investigation requirements memorandum.

New FIs were instructed to:

- read all manuals; and
- complete the home study exercises.

Home study exercises were completed electronically via the Internet before traveling to training. Exercises were graded automatically and results were posted to the CMS for FS review. Any new FI scoring less than 80 percent on the electronic home study was not allowed to attend training and was terminated from the Headway system. Based on past experience, it was evident that additional resources should not be devoted to any prospective FI unable to score at least 80 percent on the home study and that he or she should not be allowed to attend training. Appendix A contains the NTP home study memorandum, while Appendix B contains the electronic home study exercises.

4.5.2 New-to-Project Training Supplies

Using a master list of needed supplies, all supplies were prepared, ordered (if necessary), and stored in preparation for training activities throughout the survey year.

4.5.2.1 Printed Materials Related to Training

While using computers for data collection greatly reduced the production of printed materials, many paper forms were still necessary, particularly for training. A detailed, near-verbatim guide was prepared for each member of the training team. Along with the training guide, numerous printed materials were developed:

- Data Collection Agreements for all new FIs to signify they agreed to follow procedures and maintain confidentiality;
- Training Workbook that contained necessary exercises, printed examples, screening scripts, and additional instructions;
- Training Segment materials packet with example listing and locating materials for the practice segment used in training;
- Mock Scripts separately bound for four different paired mocks, including the screening and interview scripts for each case;
- Quality Control Forms specifically for the various training cases;
- Reference Date Calendars and Interview Payment Receipts for use during the practice interviews;
- Showcard Booklets, including Pill Cards, for training and use during subsequent fieldwork;
- Supplies to be used during the course of training, including the Lead Letter, Study Description, Q&A Brochure, and various tools used for obtaining participation, such as Newspaper Articles, RTI/SAMHSA Fact Sheet, Certificate of Participation, Who Uses the Data handout, "Sorry I Missed You" cards, NSDUH Highlights, and the NSDUH short reports; and

- Certification Materials used during the certification process at the conclusion of training.

4.5.2.2 Training Videos

Video segments that played directly from the trainer laptops during training provided controlled, standardized, visual presentations of the various tasks assigned to FIs. DVDs containing all video presentations were also available in case problems occurred with the videos installed on the trainer laptops. These videos contained multiple segments for use throughout the course of new FI training. Various videos detailing important screening and interviewing activities as well as administrative tasks were used in 2012. One video illustrating the MHSS recruitment process developed prior to the March 2010 training session was used at the 2012 NTP sessions. Also, minor updates were made to one video originally developed for 2006 veteran training ("Mission: NSDUH"). New FIs also viewed the two videos "Your Important Role," which is used for controlled access situations, and "NSDUH Study Results," which was updated for 2012 to include clips from the 2010 NSDUH Data Release Press Conference.

4.5.2.3 iLearning Training Program

In 2012, the electronic multimedia, interactive training application, referred to as iLearning (which stands for independent learning), continued to be used. The iLearning courses featured audio and visual training components as well as creative videos packaged onto a CD that could be viewed on the FI laptop. iLearning allowed FIs to complete training courses at their own pace and review portions of the course again as needed. Each course consisted of visual slides utilizing text and graphics, an audio component providing important information and instructions, and an assessment portion to ensure the FI's comprehension of the material presented. Upon completion of the course and transmission to RTI, the course assessment results were posted to the CMS for FS review.

The courses used during the 2012 NTP training sessions included:

- iLearning Introduction: This course provided an introduction to the iLearning program and instructions on using this and other iLearning courses.
- IRB Training: This course provided training on IRB protocols and covered the ethics and regulations involving research on human subjects.
- CIPSEA Training: This course described the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) requirements to protect information collected on NSDUH and when performing other statistical activities.
- Bilingual Training: This course was completed by NTP bilingual FIs after returning home from training.

After being in the field for 1 month, all NTP FIs were required to complete additional iLearning courses. These courses were originally developed for previous veteran training programs and included:

- Using Your Segment Materials: This course explained the overall sampling process and reviewed the proper use of the segment materials and the importance of maintaining the sample integrity. Common errors associated with using the segment materials were explained as well.
- Challenging Field Situations: This course shared approaches for handling challenging situations in the field, including controlled access, reluctant respondents, refusals, and other related topics. A brief review of the uses and importance of NSDUH data as well as excerpts from the 2008 NSDUH Press Conference video were included.
- TSG's Top 10 Equipment Tips: This course reviewed the top 10 tips from NSDUH's Technical Support Group (TSG) regarding the computer equipment. Equipment care and maintenance, troubleshooting, important reminders, computer terminology, and procedures for calling TSG staff were included.
- Habits for Success: This course was designed to review key project procedures and protocols and was completed by FIs prior to the start of each quarter. This course also addressed answering respondent questions, gaining cooperation, and other helpful refusal topics. Within the course, the assessment questions and selected content varied from quarter to quarter in order to expand the topics covered.

Creation of the iLearning courses was a complex and detailed effort, including many steps during the development and testing process to ensure all components of the course functioned properly. However, the iLearning program enabled a more individualized and interactive training model.

4.5.3 New-to-Project Bilingual Training

Interviewers who were RTI-Certified as bilingual interviewers completed the "Bilingual Training" iLearning course after returning home from training. They also completed a mock screening and interview in Spanish after completing the course.

4.6 Preparation for Veteran Field Interviewer Training

The 2012 veteran FI training program consisted of three iLearning courses completed independently at home by all veteran FIs during November and December 2011, followed by a 1-day in-person FS team meeting and training session held the first week in January 2012 at 38 sites around the country. This 38-site design allowed for smaller groups and less travel for many FIs. This section reviews the main steps necessary to prepare for this veteran training program.

4.6.1 Veteran Training iLearning and Data Collection Preparations Packages

Prior to training, all veteran FIs continuing for 2012 received a veteran training iLearning package containing:

- cover memorandum from the National Field Director, including an overview of the veteran training program, the tasks to be completed, and a detailed list of changes made to the FI Manual and Computer Manual for 2012;
- 2012 Field Interviewer Manual;

- 2012 Field Interviewer Computer Manual;
- NSDUH Materials "Keep" List outlining which 2011 materials to keep for 2012; and
- 2012 NSDUH Veteran Training iLearning CD containing the training courses to be completed prior to the FS team meeting plus one additional feedback course (to be completed after the in-person FS team meeting).

In order to prepare for training, veteran FIs were instructed to successfully complete and transmit all veteran training iLearning courses following a specified timeline.

Each iLearning course included an assessment portion with 5 to 10 questions (excluding the feedback course that was not graded) and an iLearning Feedback portion asking for feedback on the specific course. After FIs completed the iLearning courses and transmitted to RTI by the specified deadlines, the courses were scored electronically and the results were posted on the CMS. FSs reviewed any missed questions with FIs prior to their scheduled FS team meeting. Any FI not achieving a score of 80 percent or higher for each course was placed on probation and required to complete additional training before beginning Quarter 1 fieldwork. Sections 4.5.2.3 and 4.6.2.3 contain brief course descriptions.

In a separate shipment, all veteran FIs received a data collection preparations package containing:

- cover memorandum from the National Field Director;
- CD containing the 2012 CAI and iPAQ programs;
- instructions for loading the 2012 CAI and iPAQ programs; and
- veteran FI bulk supplies.

Following the completion of 2011 data collection efforts and prior to beginning Quarter 1 data collection in January 2012, veteran FIs were instructed to:

- use the provided CD to install the 2012 CAI and iPAQ programs;
- transmit a practice break-off screening and interview to RTI to confirm that the 2012 CAI and iPAQ program updates were installed correctly; and
- recycle or discard any 2011 materials not listed on the NSDUH Materials "Keep" List.

Appendix C contains the veteran training iLearning courses memorandum, and Appendix D contains the data collection preparations memorandum.

4.6.2 Veteran Training Supplies

Using a master list of needed supplies, all supplies were prepared, ordered (if necessary), and stored in preparation for training activities.

4.6.2.1 Printed Materials Related to Training

A detailed, near-verbatim Veteran Training Guide was prepared for each member of the training team. Based in part on the guide developed for 2011, most sections of the guide were newly developed to present relevant topics for 2012. Along with the training guide, other printed materials were developed:

- 2012 NSDUH Veteran Training Workbook that contained necessary exercises, printed examples, and additional instructions; and
- Attendance List/Summary Report to allow trainers to report on classroom activities and attendance at the end of the day.

4.6.2.2 Training Videos

Short videos were developed for the 2012 veteran training iLearning courses. A new video welcoming the FIs to the 2012 veteran training program was shown in the "Welcome – 2012 Veteran FI Training" introduction course.

4.6.2.3 iLearning Training Program

As explained in Section 4.5.2.3, iLearning courses were developed for the 2012 NSDUH. Refer back to Section 4.5.2.3 for additional details on the iLearning program.

The iLearning courses created and utilized during 2012 veteran training included several courses just for veteran FIs:

- Welcome – 2012 Veteran FI Training: This short introduction included a video providing an explanation of the veteran training program.
- 2012 NSDUH Updates: This training course focused on the instrumentation and material updates for 2012.
- FS Team Meeting Feedback: This course gathered feedback on the FS team meeting and was completed in January 2012 after FIs attended their FS team meeting.

Originally created for veteran training, two iLearning courses were used for both veteran and NTP training in 2012. Refer to Section 4.5.2.3 for course descriptions of these courses:

- Habits for Success; and
- CIPSEA Training.

4.7 Preparation for Field Data Collection

To prepare for data collection, a master list of needed supplies was developed. Using this list, all supplies were developed, ordered (if necessary), and stored for use in data collection activities throughout the survey year.

4.7.1 Assignment Materials

Veteran FIs were given assignment materials as each new quarter approached. These materials included a packet of segment materials (including the various maps and listing sheets for a segment) and lead letters. Letters were prepared and sent by the FIs prior to the time they would be working a particular area. Before beginning a new quarter's work, FIs also transmitted to receive their new case assignments on their iPAQs.

FIs performing well at NTP training were given assignment materials for the cases assigned to them. The assignment materials consisted only of the segment materials packet. Usually, the FS mailed the lead letters so the FI could begin work immediately upon the successful completion of training. FIs also transmitted at the end of training to pick up their assigned cases on their iPAQs. FIs struggling during training were placed on probation and received no assignments until they adequately completed further training with their FSs. Any materials for segments not assigned to an FI were sent to the FSs for later assignment.

4.7.2 Bulk Supplies

Bulk supplies were packed at RTI and shipped overnight directly to the homes of veteran FIs and new staff that completed NTP training successfully. During the year, FSs were responsible for requesting additional supplies for their FIs using a resupply ordering process on the management website. Requested items were sent from the Field Distribution Center directly to the FIs needing supplies.

4.8 Website Development

Using the influence of the Internet to broaden communication, RTI staff continued to refine and enhance the two NSDUH websites.

4.8.1 Case Management System

The up-to-date web-based CMS enhanced the ability of all levels of management to make informed decisions based on current field conditions. Each night, data were transmitted to RTI from the FIs' iPAQs and laptops for inclusion in the CMS. The next morning, each supervisor and manager had access to the results of the previous day's work and its effect on the totals for that quarter.

The CMS also contained many helpful tools such as electronic versions of the FI, FS, and RS Manuals; logs to enter new recruits and training information; links to other pertinent sites; project calendars; and other administrative tools.

Access to this secure website was tightly controlled with system-wide security provided through secure links to the network from each user's computer. Additionally, several levels of passwords were required to enter the system.

4.8.2 NSDUH Respondent Website

For computer savvy respondents, an informative public NSDUH website was maintained. Visitors to the site could access a variety of topics such as project description, confidentiality, and frequently asked questions. Brief information was included about both SAMHSA and RTI, with links to the websites of both organizations. Also included was a listing of various users of NSDUH data, which included links to those users' websites and news articles about NSDUH. Respondents could also access contact information for a NSDUH project representative via the website.

4.9 Maintaining NSDUH Equipment

Staff used an extensive inventory system to monitor the disbursement and location of all NSDUH equipment, including FI iPAQs and laptops; management laptops and printers; training projectors; and the many miscellaneous parts and cords. Technical assistance to the users of the equipment was an important and necessary task.

All field and management staff receiving NSDUH equipment acknowledged that they would not alter or add software unless directed by RTI staff to do so. Staff also indicated understanding the full and legal responsibility for taking reasonable and appropriate steps to safeguard equipment from damage, loss, or theft. All staff received training and had written manuals available explaining proper care and handling of the equipment and the consequences of repeated equipment problems.

If staff left the project, equipment was returned to Technical Support for check-in and maintenance. Detailed procedures were in place to recover any equipment not readily returned by former staff.

4.10 Problems Encountered

Development of all NSDUH materials and the computer programs for the electronic instruments requires a tight schedule in order to complete all preparations on time. There were no major problems encountered during the 2012 material preparations phase other than the compressed preparation schedule associated with implementing and testing iLearning courses.

In the fall of 2011, MHSS materials development and preparations for both FI and clinical interviewer trainings were occurring simultaneously with other normally scheduled activities, such as preparing for the 2012 veteran and NTP FI trainings. With limited time for implementation, RTI staff made the necessary revisions to the instruments, manuals, and materials so that data collection for both the main study and the MHSS could begin as scheduled in January 2012.

Exhibit 4.1 2012 iPAQ Updates

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH SCREENING APPLICATION UPDATES

The following updates were made to the 2012 NSDUH Screening Application:

- No updates were made to the screening program other than updating dates.

Exhibit 4.2 2012 CAI Changes

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH CAI INSTRUMENT REVISIONS

Module Specific

Introduction

- Updated the computer-assisted interviewing (CAI) instrument version.

Tobacco

- Added words "tobacco in" to CGREF5.

Health Insurance and Income

- Updated state program names for Medicaid, Children's Health Insurance Plan (CHIP) and Temporary Assistance for Needy Families (TANF).

References

Center for Behavioral Health Statistics and Quality. (2011a, September). Illicit drug use among older adults. *The NSDUH Report*. [Available at http://www.samhsa.gov/data/Short_Reports.aspx]

Center for Behavioral Health Statistics and Quality. (2011b, February). Young alcohol users often get alcohol from family or home. *Data Spotlight*. [Available at http://www.samhsa.gov/data/Spotlights.aspx]

Office of Applied Studies. (2010, March). 12 year olds more likely to use inhalants than cigarettes or marijuana. *OAS Data Spotlight*. [Available at http://www.samhsa.gov/data/Spotlights.aspx]

5. Field Staff Training

Training for all levels of project field staff occurred both prior to the start of data collection and throughout the year. Having experienced staff enabled training programs to focus on enhancing and improving necessary project skills rather than simply teaching the basic steps.

5.1 Management Training Programs

With a highly experienced management team, there was no formal management session conducted in preparation for the 2012 National Survey on Drug Use and Health (NSDUH). In order to share important topics about the 2012 NSDUH, field management staff completed two veteran training iLearning courses in November 2011. The courses included:

- Habits for Success; and
- 2012 NSDUH Updates.

The course details for the first course are provided in Section 4.5.2.3, while details for the last course are provided in Section 4.6.2.3.

During the course of 2012 data collection, a management meeting was held on August 7, 2012, in Raleigh, North Carolina, to share field management techniques and strategies for success. The meeting also served to reenergize field management staff and build a greater sense of cohesiveness for the geographically dispersed team. Regional directors (RDs), regional supervisors (RSs), field supervisors (FSs), RTI survey specialists, the National Field Director, the associate project director, and several staff from the Substance Abuse and Mental Health Services Administration (SAMHSA) attended the educational and interactive sessions.

General topics covered during the meeting included:

- maintaining data integrity; and
- managing project resources

Following the meeting, notes taken during the sessions were compiled, summarized, and posted to the Case Management System (CMS) for reference by all project management staff.

5.2 New-to-Project Field Interviewer Training Sessions

5.2.1 Design

Training sessions were held prior to the start of each new quarter throughout the year to train newly hired new-to-project (NTP) field interviewers (FIs). These sessions helped maintain a sufficient staff size to complete screening and interviewing within the quarterly time frames. For each session, there were multiple training rooms staffed by teams of four trainers. Occurring January 20–January 26, March 23–March 29, June 22–June 28, and September 21–September 27, a total of 196 new FIs were trained during these replacement sessions. The January NTP

training session was held in Baltimore, Maryland. The remaining sessions were held in Cincinnati, Ohio. At the end of this chapter, Table 5.1 summarizes the interviewer training sessions conducted for the 2012 NSDUH.

The NTP training program consisted of 7 days of training covering general techniques of interviewing, screening using the iPAQ handheld computer, procedures for conducting NSDUH interviews on the laptop computer, general NSDUH protocols, and technical support. After returning home from training, Spanish-speaking FIs completed a "Bilingual Training" iLearning course and an individual mock screening and interview to review the Spanish translations of the questionnaire and the iPAQ screening program.

All new FIs were required to pass an individually conducted certification process as part of the successful completion of training. Each FI had to demonstrate knowledge of the basic NSDUH protocols by completing a straightforward screening and interview with an abbreviated version of the audio computer-assisted self-interviewing (ACASI) questions.

Of the 196 new FIs trained during 2012, 3 FIs were placed on probation based on their overall performance during training and on problems with the certification process. An additional six FIs were released during training due to their inability to meet training expectations. Another FI resigned during training due to health issues. One FI was unable to complete training due to illness and resigned before an in-person training of the missed content could be conducted.

To provide consistency between training classrooms, a near-verbatim guide with 23 sections provided detailed instructions and text to ensure all necessary instructional points were covered. In addition to the guide, trainers also used 12 videos (saved to the trainer laptop as well as on a set of six DVDs for backup) that contained multiple video segments for use throughout training, an iLearning CD containing courses used throughout training and after training, a workbook containing exercises on the iPAQ and laptop computer and printed examples, training segment materials used in exercises that replicated actual segment materials, the FI Manuals for reference, and the two computers (the iPAQ and the laptop) with accessory equipment.

5.2.2 Staffing

At each training site, staff included a site leader, a logistical assistant, a lead technician, a certification coordinator, and one or more training teams. Each of these roles was well defined to ensure that training progressed smoothly.

The site leader at each training site coordinated all FI registration activities, hotel relations, and logistics and monitored FIs and trainers. The site leader's specific tasks included:

- overseeing the fingerprinting process of new FIs;

- coordinating all services provided by the hotel with the assigned hotel representative;
- managing the trainers and training rooms;
- evaluating FI performance and working with trainers to resolve problems with FIs, including probation or even termination when necessary as a last resort;
- reporting the status of training to management and supervisory staff each evening using the provided Daily FI Training Evaluation (see Exhibit 5.1 at the end of this chapter);
- supervising the certification process and making final decisions about the status of any FIs failing ; and
- informing trainers about resolutions to any questions, problems, or suggestions following consultation with appropriate project staff.

The site leader role was filled by a retired NSDUH RD who was contracted to be the site leader at all NTP training sessions and who had extensive experience with project protocols and management goals.

The logistical assistant worked closely with the site leader throughout training to ensure all FIs were registered properly, all training rooms had all necessary supplies, and hotel services functioned smoothly. Other duties included grading in-class assignments and distributing training and incentive checks at the successful conclusion of training.

The lead technician served as the point of contact for all technical issues including the proper functioning of all equipment and programs. Other duties included supervising training equipment setup and the initialization and distribution of FI computer equipment.

The certification coordinator managed the certification process, including establishing appointment schedules, monitoring and distributing certification supplies and materials, and reporting the results to the site leader. In an effort to reduce costs, the logistical assistant also served as the certification coordinator at all of the NTP training sessions.

Each classroom was taught by a training team consisting of a lead trainer, two assistant trainers, and a technical support representative. The lead trainer and assistant trainers divided the responsibility for presenting most sections of the training, while the technical support representative presented portions of the equipment-related sections. The lead trainer had the additional responsibility for the logistics and schedule of the training room. In general, one trainer would train from the front of the room while the other trainer(s) would monitor FI progress, assist FIs with questions, and sometimes operate the computer equipment.

In addition to training the equipment-related sections, the technical support representative prepared and set up the computers for each FI; ensured the proper functioning of the iPAQ, laptop, and projection equipment used for the training presentation; and provided in-class technical help.

Training teams were selected based on availability and experience. The lead trainer was usually an RS with considerable training experience or an experienced instrumentation or

operations team member. Assistant trainers were usually RSs, FSs, instrumentation team members, or survey specialists.

5.2.3 Content of New-to-Project Field Interviewer Training Sessions

5.2.3.1 Day 1

After completing the registration process the evening before, training classes began first thing in the morning with an introduction to the history and scope of NSDUH presented in a video featuring Project Director . FIs also became familiar with the project organization via a creative video titled, "Mission: NSDUH." Next, FIs reviewed the FI job description and responsibilities and then completed an introductory computer session. This session included instruction in the use of the laptop computer hardware and the basics of the iPAQ hardware and software, although the actual screening program was not covered. For much of the afternoon, FIs learned how to contact selected households and how important it was to be knowledgeable about the study. They also discussed professional ethics and respondents' rights, and they had the opportunity to review supplementary materials and practice making effective introductions and answering respondent questions. At the end of the day, FIs were introduced to iLearning, a multimedia, computerized training program. On the evening of Day 1, FIs used iLearning to complete the "IRB Training" course, which covered ethics and regulations involving research on human subjects, the role of the Institutional Review Board (IRB), and the role of the interviewer in protecting respondents' rights. They also completed the "CIPSEA Training" iLearning course, which described the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) requirements to protect information collected on NSDUH and the role of the Office of Management and Budget (OMB) in providing oversight and designating statistical agencies under CIPSEA. All FIs were invited to attend an evening field interviewer lab (FI Lab) session for additional practice or assistance with the iLearning homework; FIs with little computer experience were especially encouraged to attend the FI Lab for hands-on practice in order to build their confidence.

5.2.3.2 Day 2

Day 2 included a general introduction to survey sampling and counting and listing, followed by an in-depth discussion of how to locate segments and selected dwelling units (DUs). Trainers then introduced the screening process using a video of a mock screening. Following a trainer demonstration, each FI had the opportunity to operate the iPAQ during a group walk-through screening exercise. Discussions on quality control and record of calls were also included in the afternoon. FIs had the opportunity to operate the iPAQ again during a group screening exercise conducted via round-robin. The training day ended with a discussion of screening and interviewing result codes. All FIs were again invited to attend an evening FI Lab for additional practice.

5.2.3.3 Day 3

On Day 3, FIs focused on gaining experience and confidence by conducting numerous practice screenings on the iPAQ. They completed an enumeration and rostering exercise round-robin style, as well as individual and paired mock exercises covering the entire screening

process. FIs also learned about screening refusal codes and refusal reports. Next, trainers presented a brief discussion of the functions of the computer-assisted interviewing (CAI) manager program on the laptop, including an overview of the NSDUH e-mail system. FIs were then introduced to the NSDUH interview and the basics of good field interviewing techniques. A video of a mock interview provided an overview of the process. This was followed by discussions on bias and probing, as well as the importance of following conventions. FIs then practiced transitioning from the screening to the interview with a partner. All FIs were again invited to attend an evening FI Lab for additional practice.

5.2.3.4 Day 4

On Day 4, FIs learned the details of the NSDUH interview with a round-robin read-through of the entire questionnaire, including question-by-question specifications. This was followed by a discussion of the details required in collecting industry and occupation information. During a paired exercise, FIs practiced administering the industry and occupation questions and probing with a partner. Next, FIs completed an individual practice interview exercise that allowed them to review both the format and questions in the CAI program at their own pace. Lastly, FIs were able to provide feedback on the training session by completing a brief questionnaire at the end of Day 4. Interested FIs could attend an FI Lab in the evening.

5.2.3.5 Day 5

FIs began Day 5 with a session on transmitting data. A trainer demonstrated how to transmit from both the iPAQ and the laptop. The class then began a series of paired mock exercises encompassing the entire screening and interviewing process so FIs could practice the transition from the screening on the iPAQ to the CAI questionnaire on the laptop. Following the mock interviews, a group review was conducted by the trainer. At some point during the practice mock interviews, FIs completed a successful transmission on both computers with assistance from the technical support representative. Classes then discussed the important topic of dealing with reluctant respondents and overcoming other difficult situations. This session included informative video segments and group exercises. All FIs were given the option of attending an evening FI Lab.

FIs who were performing well could attempt the certification process the evening of Day 5.

5.2.3.6 Day 6

Training on Day 6 began with an explanation on the specifics of screening a group quarters unit (GQU), followed by details on checking for and adding missed DUs. FIs then learned about other screening topics that they may encounter during their work, but not on a regular basis. Topics included editing addresses, placing cases on hold, re-opening cases, and using optional iPAQ accessory programs for organization and planning purposes. After lunch, FIs had the opportunity to complete another series of paired mock exercises to further practice the entire screening and interviewing process. Following the mock interviews, a group review

was conducted by the trainer. Certifications and an FI Lab were scheduled for the evening of Day 6.

5.2.3.7 Day 7

Day 7 included an explanation of the Mental Health Surveillance Study (MHSS), an additional research study conducted during 2012, and the FI's role on that study. This section included a video demonstrating the MHSS recruitment process. After viewing the video, FIs completed a mock exercise in order to practice the recruitment process. This section was followed by discussions on administrative procedures and proper documentation and reporting. The next section on troubleshooting and technical support informed FIs about the most common technical problems they might encounter and the steps to correct them, as well as when and how to contact technical support for additional help. Care and maintenance of the computer equipment was also discussed. A brief recap of the entire process of screening and interviewing helped FIs review how all the tasks fit together. This recap included discussions on FI performance, adherence to procedures and quality control, and the NSDUH study results. FIs then completed a final evaluation to provide feedback on the completed training session. Any remaining certifications took place at the conclusion of Day 7.

5.2.4 New-to-Project Bilingual Training

Following training, bilingual FIs completed a 1-hour "Bilingual Training" iLearning course on the Spanish-language NSDUH materials and questionnaires. This course reviewed the Spanish versions of the iPAQ screening program, NSDUH interview, and other 2012 supplemental materials, as well as the differences from the English versions. Only those FIs who had been hired as bilingual FIs completed this iLearning course. Following successful completion of the course and an individual mock screening and interview to review the Spanish version of the questionnaires in detail, bilingual FIs were deemed RTI-Certified and, as such, are the only FIs allowed to conduct the NSDUH interview in Spanish.

5.2.5 Mentoring of New-to-Project Graduates

After completing the NTP training program, all graduates were mentored in the field by an experienced FI, their FS, or another FS. Mentoring of all new FIs was required and usually occurred within a week following the conclusion of training, preferably during the graduate's first trip to the field.

Mentors were given standardized instructions (see Exhibit 5.2) to be sure all important protocols learned during training were reinforced.

5.2.6 New-to-Project Post-Training iLearning

After 1 month in the field, NTP FIs were required to complete three additional iLearning courses—Using Your Segment Materials, Challenging Field Situations, and TSG's Top 10 Equipment Tips. Before beginning each subsequent quarter of work in 2012, NTP FIs also completed the "Habits for Success" iLearning course as described in Section 5.5. Refer to Tables 5.3 and 5.5 for the results of these courses.

5.3 Veteran Field Interviewer Training Sessions

5.3.1 Design

To prepare the FIs chosen to continue working from the 2011 NSDUH into 2012, the veteran FI training program consisted of three pre-training iLearning courses (see Section 4.6.1 for more information and Tables 5.4 and 5.5 for the results of these courses) completed in November and December 2011, followed by a 1-day FS team meeting and training session held in January 2012 at 38 different sites across the United States. Having regional sessions throughout the country served several purposes:

- Through the developed training program, project management staff expressed appreciation for past efforts and provided explicit instructions for ways to improve future performance.
- FIs were able to share helpful tips with each other.
- FSs met with their entire team to discuss specific issues for their assigned area and enhance team rapport.

FS team meetings were held at 38 sites listed in Table 5.2. Two separate sessions were held, with the A groups meeting January 3 and the B groups meeting January 5, 2012. In addition to these early January FS team meetings, two make-up teleconferences were held on January 5 and 9, 2012, to train any veteran FIs unable to attend their FS team meeting. Also, throughout 2012, additional veteran FIs who missed the January sessions were trained with permission on an individual basis. Table 5.1 summarizes the January veteran training sessions.

The 1-day (7-hour) FS team meeting and training session covered topics such as region-specific goals, FS team results, and MHSS procedures and included four workshops focusing on techniques for gaining cooperation among difficult respondent populations and averting refusals, production cost and quality, controlled access and difficult-to-contact cases, and an FI roundtable of various topics. All FSs presented the workshop on gaining cooperation and averting refusals and then were able to choose one of the remaining three workshops to present to their team. The shortened training session was possible due to the minimal changes made to the screening and interviewing programs and allowed many FIs to avoid an overnight stay.

To provide consistency between FS team meetings, a training guide with seven sections provided detailed instructions and text to ensure all necessary instructional points were covered. The FS team meeting approach allowed the FS to choose from three different available workshop topics that were scripted but still allowed individual FS ideas to be included in the workshop.

5.3.2 Staffing

At each training site, there was a site leader who served as the contact person for any hotel or logistical issues, served as the liaison between the site and NSDUH management, and ensured all classrooms followed the guide and remained on schedule. In addition, assistant trainers were available at larger sites to assist with classroom setup and training activities as needed. The assistant trainer was an extra NSDUH staff trainer at the site.

Each class was taught by the managing FS, assisted by the site leader and/or assistant trainer. Each FS team attended one session, either Session A or Session B. One trainer usually presented at the front of the room while the other trainers monitored FI progress and assisted FIs with questions and activities.

5.3.3 Train-the-Trainers

To prepare FSs, site leaders, and assistant trainers for their training role and to instruct all project staff in the changes for the 2012 survey, a Train-the-Trainers session was held via a video-streamed meeting. Staff were able to view a video presentation of the meeting in real time from their computer and submit any questions through a secure network. The 2-hour session was held on November 13, 2011. The video-streamed meeting was led by members of the instrumentation team who reviewed all portions of the veteran training guide and materials as well as logistics for the January sessions.

To further prepare staff for the upcoming team meetings, each RS team had a group conference call in December to review the veteran training guide in more detail. Each FS was assigned a section of the guide to review and then share information about the content, detailed preparations, and proper handling of any exercises and issues that might arise. These calls allowed the FSs, under the leadership of the RS, to review the guide as a small group and better prepare to be trainers at the FS team meetings. An instrumentation team member was on each RS call to help answer any questions about the guide.

5.3.4 Content of Veteran Field Interviewer Training Sessions

The 1-day FS team meeting and training session began at 9:00 am with an FS welcome to the session, an ice-breaker activity, and a review of the 2011 FS team results, training agenda, and goals. Next, FIs were introduced to administrative and timekeeping changes related to the Service Contract Act (SCA). FIs then were trained on new transmission software for the FI laptop to be released in Quarter 1, 2012. The next workshop focused on MHSS procedures and materials. After lunch, FSs conducted a workshop that included practice exercises on answering respondent questions, a brief summary of selected 2010 NSDUH results and how to use that information with respondents, a discussion on gaining cooperation among difficult respondent populations, answering respondent questions accurately, and techniques on refusal aversion and conversion. FSs could tailor the workshop discussion to specific problems faced by their region. FSs then had the choice of conducting one of three interactive workshops they felt were most suited to the issues faced in their regions, including production cost and quality, controlled access and difficult-to-contact cases, and an FI roundtable. The day ended at 4:00 pm with a wrap-up in which FI questions could be answered.

5.3.5 Special Veteran Training Sessions

Two additional make-up teleconferences were held on January 5 and 9, 2012, to accommodate those veteran FIs unable to attend the early January sessions. An instrumentation team member served as the trainer for these sessions so that FSs could focus on managing data collection.

As the year progressed, veteran FIs from 2011 who wished to resume working were trained individually via iLearning and telephone conference with an FS or an instrumentation team member. These FIs missed the January sessions due to illness or preapproved scheduling conflicts. With special permission, one-on-one training brought these FIs up to speed on the 2012 NSDUH. Following successful completion of the iLearning courses, the FS or instrumentation team member worked with the veteran FI(s) to cover the content of the 2012 FS team meeting and training session.

5.4 Ongoing Training

Regional team meetings with particular FS teams occurred throughout the year. As needed, team meetings were held to introduce FIs to a new supervisor (either FS or RS). In other situations with teams performing below expectations, the focus of these meetings was to provide further training for FIs on refusal avoidance, refusal conversion, and efficiently working case assignments. Additional discussion topics included data quality and specific team performance issues. Three of these in-person team meetings occurred during 2012 for FI teams in Maryland, Vermont, and West Virginia. These meetings were attended by the team's FS, RS, and RD.

5.5 Periodic Evaluations

Periodic evaluations of interviewer knowledge were conducted via the "Habits for Success" iLearning course completed prior to each quarter of work in 2012. This iLearning course was available to FIs via the iLearning CD they received as part of the veteran or NTP training programs. FIs had 1 week to successfully complete the "Habits for Success" course, which covered basic NSDUH protocols and procedures, answer 10 assessment questions, and transmit the course to RTI by the specified deadline, approximately 2 weeks before the end of each quarter. The assessment portion of the course was then scored after being transmitted to RTI. Each quarter, FIs reviewed the "Habits for Success" iLearning course, which contained selected content that varied each quarter and other topics that remained constant. However, the set of 10 assessment questions at the end of the course changed completely each quarter to test interviewer knowledge of basic NSDUH protocols and avoid repetition. To pass the course, FIs had to score at least 80 percent. FIs not achieving a passing score were placed on probation and required to complete retraining with their FS prior to beginning work the next quarter. See Table 5.5 for the results of the "Habits for Success" iLearning course.

5.6 Problems Encountered

Leading the training sessions held throughout the year required involvement of project staff with other NSDUH responsibilities. These dedicated staff trained each day and then completed their other project duties in the evenings. The demands on trainer time were increased

on evenings when they had to staff FI Labs or conduct certifications. Training planners tried to rotate staff across the various training assignments throughout the year to avoid overloading any one individual. This approach worked well. Planning for the 2012 veteran training sessions also required extensive involvement of project staff with other ongoing NSDUH responsibilities to establish contracts and coordinate the logistics with 38 different training locations. Staff worked diligently to ensure the contracts and training arrangements with various hotels across the country were in place in time for the January 2012 FS team meetings.

Table 5.1 2012 NSDUH FI Training Programs

Month	FI Training Sessions *Date and Location*	FIs Trained	Cumulative Number of FIs Trained	Attrited FIs	Cumulative Number of Attrited FIs
January	***Veteran Training Sessions*** Dates: *Session A*: 1/3 *Session B*: 1/5 Location: 38 sites (see Section 5.3.1)	615	615	4	4
	Makeup Veteran Trainings Date: 1/5 & 1/9 Location: Teleconference	25	640	0	4
	Veteran FIs trained one-on-one	3	640	0	4
January	***New-to-Project Training Session*** Date: 1/20-1/26 Location: Baltimore (MD)	55	695	0	4
February	No training session	0	695	7	11
March	***New-to-Project Training Session*** Date: 3/23-3/29 Location: Cincinnati (OH)	56	751	7	18
April	No training session	0	751	15	33
May	No training session	0	751	18	51
June	***New-to-Project Training Session*** Date: 6/22-6/28 Location: Cincinnati (OH)	36	787	12	63
July	No training session	0	787	17	80
August	No training session	0	787	12	92
September	***New-to-Project Training Session*** Date: 9/21-9/27 Location: Cincinnati (OH)	49	836	17	109
October	No training session	0	836	26	135
November	No training session	0	836	10	145
December	No training session	0	836	17	162

FI = field interviewer.

Table 5.2 2012 NSDUH Veteran Training Sites

Date	Locations		
Session A: January 3, 2012	Akron, OH Albuquerque, NM Atlanta, GA Boise, ID Charleston, WV Charlotte, NC Chicago, IL Denver, CO Des Moines, IA Detroit, MI Fort Smith, AR Honolulu, HI Houston, TX	Los Angeles, CA Louisville, KY Minneapolis, MN Naperville, IL Nashville, TN New Orleans, LA Newton, MA Ocala, FL Philadelphia, PA Pittsburgh, PA Portland, OR Providence, RI Queens, NY	Richmond, VA Rockville, MD Salt Lake City, UT Syracuse, NY West Palm Beach, FL
Session B: January 5, 2012	Anchorage, AK Cincinnati, OH Dallas, TX Grand Rapids, MI	Las Vegas, NV Phoenix, AZ Portland, ME	

Table 5.3 Results from New-to-Project Post-Training iLearning

	Passed		Failed		
Course Name	**Count**	**%**	**Count**	**%**	**Total**
Using Your Segment Materials	186	100.0	0	0.0	186
Challenging Field Situations	184	98.9	2	1.0	186
TSG's Top 10 Equipment Tips	185	100.0	0	0.0	185
Bilingual Training (Bilingual New-to-Project FIs ONLY)	37	100.0	0	0.0	37

FI = field interviewer, TSG = Technical Support Group.

Table 5.4 Results from Veteran Training iLearning Courses

	Passed		Failed		
Course Name	**Count**	**%**	**Count**	**%**	**Total**
2012 NSDUH Updates	649	99.8	1	0.2	650
CIPSEA Training	650	100.0	0	0.0	650

CIPSEA = Confidential Information Protection and Statistical Efficiency Act.

Note: The difference in the number of field interviewers (FIs) completing the courses is due to FI turnover.

Table 5.5 Results from Periodic iLearning Evaluations

Course Name	Passed		Failed		Total
	Count	%	Count	%	
Q1 2012 Habits for Success	649	99.7	2	0.3	651
Q2 2012 Habits for Success	679	100.0	0	0.0	679
Q3 2012 Habits for Success	693	100.0	0	0.0	693
Q4 2012 Habits for Success	687	100.0	0	0.0	687

Q = quarter.

Note: The difference in the number of field interviewers (FIs) completing the courses is due to FI turnover.

Exhibit 5.1 Daily FI Training Evaluation

Exhibit has been removed.

Exhibit 5.1 Daily FI Training Evaluation (continued)

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions (continued)

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions (continued)

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions (continued)

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions (continued)

Exhibit has been removed.

Exhibit 5.2 Mentoring Instructions (continued)

Exhibit has been removed.

6. Data Collection

This chapter presents the basic data collection procedures provided to field staff working on the 2012 National Survey on Drug Use and Health (NSDUH). For further details or specific instructions, consult the 2012 NSDUH Field Interviewer Manual.

6.1 Contacting Dwelling Units

Field interviewers (FIs) were assigned specific sample dwelling units (SDUs) to contact, with the addresses or unit and location descriptions displayed on the Hewlett-Packard iPAQ handheld computer. The sample was released in partitions, with additional units made available as needed, depending on progress made during the initial weeks of data collection each quarter.

6.1.1 Lead Letter

Initial contact with residents of the specific SDUs was made through a lead letter that gave a brief explanation of the nature of the study and its methods. The letter was printed on United States Public Health Service/Department of Health and Human Services (DHHS) letterhead and signed by both the Substance Abuse and Mental Health Services Administration (SAMHSA) National Study Director and the RTI National Field Director.

For all housing units with a complete address (i.e., not a location description), letters preprinted with the addresses were included with the assignment materials distributed to FIs each quarter. FIs reviewed all addresses to check that they could be mailed, signed the letters, and mailed them via first class mail prior to and throughout the first part of the quarter so that the letters arrived fairly close to the time the FI expected to be in the area. Group quarters units and any housing units lacking a complete mailing address were not sent a letter. To allow for these cases and other instances of delivery problems, each FI had additional letters to give to respondents during a personal visit. A copy of the letter, in both English and Spanish, was also included in the Showcard Booklet for reference.

6.1.2 Initial Approach

Before knocking on the door of an SDU, the FI selected the appropriate case for that specific unit on the iPAQ. Each FI possessed a personalized letter of authorization printed on SAMHSA/DHHS letterhead authorizing him or her to work on the study and approached the door of the SDU with his or her RTI identification badge clearly visible. The FI also carried a variety of informative materials such as Question and Answer Brochures, NSDUH Highlights, and copies of newspaper articles about NSDUH.

6.1.3 Introduction, Study Description, and Informed Consent

When contacting the unit, the FI asked to speak with an adult resident (18 or older) of the unit who could serve as the screening respondent. The FI introduced himself or herself and the study. As scripted on the iPAQ, the FI mentioned the lead letter and, on the Informed Consent

screen, read the informed consent text to the screening respondent and gave him or her a copy of the Study Description. The Study Description, which was also included in the Showcard Booklet for reference, explained the purpose and sponsor of the data collection effort, assured the respondent that all information gathered would be handled in the strictest confidence, and estimated the time required to complete the screening and interview. The Study Description also stated that respondents were free to withdraw from the study at any time. Providing the Study Description and reading the scripted informed consent text from the iPAQ fulfilled all required aspects of Informed Consent for the screening portion of the study.[1]

6.1.4 Callbacks

If no respondent was available or another situation was found at the unit so that screening could not be completed during the first visit, a minimum of four callbacks were made to the unit so that each SDU was visited at least five times in an effort to complete the screening. These contacts were made at different hours on different days of the week to increase the likelihood of completing the screening. The only exception to this protocol was in case of adamant refusals. Refer to Section 6.7 for details on refusal conversion procedures.

Screening cases that had received the initial visit plus at least four callback attempts were eligible for finalization with no additional fieldwork. However, before finalizing a case, field supervisors (FSs) reviewed the iPAQ Record of Calls (ROC) of pending screenings to ensure each case had been given ample opportunity to be completed. If feasible and cost-effective, additional callbacks were made to SDUs that were not visited at different times on certain days. If the screening was not completed during these additional contacts, then the case was assigned a final code.

6.2 Dwelling Unit Screening

Screening was performed at each SDU by obtaining information about the residents of the unit to determine whether any household member would be eligible for the NSDUH interview based on the ages of the SDU members. The screening program guided the FIs through the process of asking age, gender, race/ethnicity, and military status for all persons aged 12 or older who lived at the unit for most of the calendar quarter, and the information was entered into the iPAQ.

6.3 Within-Dwelling Unit Selection

Once the roster information was entered and verified, the FI started the within-dwelling unit selection algorithm on the iPAQ. The iPAQ automatically determined, based on the composition of the household roster, whether or not anyone in the unit was selected for the interview.

---

[1] Since RTI began conducting this survey, there have been no reported incidents involving a breach in confidentiality or any problems as a result of respondents' participation in the survey. Based on that information, RTI's Institutional Review Board determined that participation in NSDUH does not pose any known risk to its participants. Therefore, the standard "no known risks or benefits" phrase is not required as part of the informed consent process.

The algorithm allowed for the selection of zero, one, or two members of a household for an interview. Dwelling units with 12- to 17-year-olds on the roster were more likely to have persons selected for an interview.

In order to identify each selected individual, the iPAQ displayed the person's roster number (based on the order in which household members were listed), the age, gender, race/ethnicity, and either the relationship to the householder (for housing units) or a first name (for group quarters units). Also listed on the iPAQ was a QuestID number, which was required to start the computerized interview on the laptop. FIs transmitted all completed screening data contained on the iPAQ to RTI each evening.

6.4 Interview Administration

6.4.1 Informed Consent and Getting Started

Once the selected individual(s) was identified during screening, the FI asked to complete the interview(s) during that visit. If this was not convenient for the respondent, the FI entered information about possible times for future contacts in the iPAQ ROC. A minimum of four visits was made at different times of day on different days of the week in an attempt to complete the interview.

Once a State reached the contractual weighted overall response rate target for the year (65 percent), achieved an 80 percent unweighted interview response rate, and completed the minimum cumulative number of interviews, interview cases that had received at least four callback attempts were eligible for finalization with no additional fieldwork. Before finalizing a case, FSs reviewed the ROC of pending interviews to ensure each case had been given ample opportunity to be completed. If feasible and cost-effective, additional callbacks were made to SDUs that were not visited at different times on certain days. If the interview was not completed during these additional contacts, then the FI assigned a final code.

For adults selected for the NSDUH interview, the FI read the Introduction and Informed Consent for Interview Respondents Age 18+ script from the Showcard Booklet to introduce the study, describe the interview process and procedures to be followed, and detail the number of people each respondent represented. Along with reading the Informed Consent script, the Study Description was also provided to meet the Informed Consent requirements for the interview. After receiving consent, the FI began the interview in a private location.

If the selected individual was aged 12 to 17, the FI was responsible for obtaining verbal consent from a parent or guardian before contacting the youth. The only exceptions to this rule were in certain group quarters situations, like dormitories, and other SDUs where consent was unobtainable because a youth was living independently without a parent or guardian residing in the home. Beginning in October 2012, this exception only applied to 17-year-olds living independently. For all youths aged 16 years or younger, parental permission was required with no exceptions.

In the Showcard Booklet, separate text for parents and guardians was included in the Introduction and Informed Consent for Interview Respondents Age 12-17 script. Once parental

permission was granted, the FI approached the youth and read the Introduction and Informed Consent script to introduce the study, describe the interview process and procedures to be followed, and detail the number of youths each respondent represented. The FI also provided a copy of the Study Description to fulfill all required aspects of Informed Consent. After obtaining the youth's agreement to participate, parents were then asked to leave the interview setting to ensure the confidentiality of the youth's responses and the FI began the interview.

6.4.2 Computer-Assisted Interviews

The NSDUH interview began in the computer-assisted personal interviewing (CAPI) mode, with the FI reading the questions from the computer screen and entering the respondent's replies into the computer. After completing the Reference Date Calendar, the FI explained to the respondent how to use the computer for the audio computer-assisted self-interviewing (ACASI) sections. Utilizing ACASI methodology for the sensitive substance use and nonuse questions enhanced privacy since the respondent listened to the prerecorded questions through the headphones and entered their responses directly into the computer. Beginning with a practice session, which introduced the various computer keys used during the interview, the respondent then proceeded through the interview. Four times during the ACASI portion of the interview, the respondent was instructed to ask the FI for a specific picture pill card designed to aid respondent recall. When the respondent was finished with the ACASI portion, the FI once again took charge of the computer, asking additional demographic questions as well as health care, insurance, and income questions. During both the beginning and ending CAPI portions, showcards were utilized to assist respondents in answering the questions.

The average computer-assisted interviewing (CAI) administration times are provided in Tables 6.1 through 6.35 at the end of this chapter for the overall survey and for the various sections of the NSDUH interview by respondent age (youths aged 12 to 17 or adults aged 18 or older) and survey year (2010, 2011, and 2012). These timing tables were calculated using audit trail data, which records responses and the time spent on each item. Cases with extreme values for the overall time (less than 30 minutes or more than 240 minutes) are excluded from the tables.

Please note that the total number of interviews included varies between tables due to interview skip patterns and excluded and missing timing data. Also note that variations in the questionnaire content between the survey years (e.g., questions added or deleted) may affect the comparability of some timing statistics.

6.4.3 End of Interview Procedures

After the last interview question, the interview process involved several final steps. FIs had to:

- prepare the Quality Control Form and ask the respondent to complete the remaining items on the form;
- have the respondent seal the completed Quality Control Form in a postage-paid envelope addressed to RTI;
- give the respondent the $30 cash incentive;

- prepare the Interview Payment Receipt, giving the appropriate copy to the respondent;
- provide the adult respondent or parent or guardian of a youth respondent with a Question and Answer Brochure if not provided earlier;
- complete the FI Observation Questions;
- enter the final result code in the iPAQ;
- gather all interview equipment and materials; and
- thank the respondent.

Each week, FIs sent all completed Reference Date Calendars and Interview Payment Receipts to his or her FS. FIs mailed sealed Quality Control Form envelopes to RTI within 24 hours of completing the interview. Each night FIs transmitted interview data to RTI.

6.5 Data Collection Management

Project management on this massive study can be summed up in one word: *communication*. For instance, the following project management meetings were conducted each week via teleconference:

- FIs throughout the country reported to their FS at least once each week to discuss production, problems encountered and possible resolutions, feedback on past work, plans for the next week, and any administrative issues.
- FSs reported to their regional supervisor (RS) weekly, discussing production, costs (including cost containment issues), goals, staffing, and other administrative issues.
- Each regional director (RD) held a weekly meeting with his or her staff of RSs to share project news and goals while addressing any problems within the region. If a particular topic needed special attention during the survey year, the RDs conducted group calls with all their RSs and FSs.
- All RDs met each week with the National Field Director and the project director.
- All directors and other key management staff met weekly with SAMHSA representatives.

Although the more formal meetings were held weekly, staff communicated almost constantly through the use of e-mail. This increased awareness of project issues by effectively passing information through the various management levels. The capability to send messages to FIs using a one-way electronic messaging system on their project laptop computer allowed for timely sharing of information with all field staff.

With the web-based project Case Management System (CMS), all management staff had access to a tremendous amount of information on the status of events in the field. Additional details on the CMS are provided in Section 8.2.

If an FS was not meeting project expectations in the FS role itself, a Performance Improvement Plan was developed. This plan clearly stated the problems noted and the steps the FS should take to improve his or her job performance. Both the RS and RD were involved in

developing the Performance Improvement Plan. In addition, RSs and RDs conducted a series of in-person training sessions with FSs to address performance concerns.

6.6 Controlled Access Procedures

At times during the data collection process, FIs had difficulty gaining access to particular SDUs. FIs with challenging circumstances were instructed to be observant, resourceful, and keep their supervisors informed of the situation. Additional suggestions taken from FS experience or from the "Guide to Controlled Access Situations" were discussed. Conversations with managers and owners generally centered on the importance of the study, SAMHSA and RTI's emphasis on confidentiality, and an individual's right to make a personal decision about participation. Supervisors sometimes contacted managers and owners directly to answer questions or concerns.

Due to prior efforts by staff who listed the dwelling units, many access problems were resolved readily. Listers recorded contact information and other steps followed to secure access so that FIs could follow the same strategies or build on already-established relations. Supervisors at the listing stage used special reports on the CMS to monitor access situations; supervisors for screening and interviewing used the same reports and recorded additional information to update the reports.

A Doorperson Letter was available to FIs to use during their work in the field. FIs carried this letter to support or supplement conversations with doorpersons, guards, and building representatives. The letter was not used with other individuals or respondents.

For continuing problems, RTI had a system to generate individualized letters and packets of information about the project. When required, FSs provided information to RSs, who then requested the packets. Upon receiving the request, specialists prepared a cover letter and assembled materials to fit the situation. The packet was sent via overnight express delivery to increase the importance placed on the contents and ensure timely delivery. A video that further explained the need for access was also available for inclusion in the packets. For situations involving university or military housing, an Institutional Review Board summary was included in the packet. Packets for situations involving university housing also received a letter of endorsement from the presidents of Duke University and the University of North Carolina.

A Law Enforcement packet could be sent to local police departments or other government agencies prior to starting data collection or after receiving a request for more information from a law enforcement official. This packet informed local law enforcement about the NSDUH and encouraged cooperation and the dissemination of information about the study to appropriate personnel. Similar to other individualized packets, it included an informative letter addressed to an appropriate recipient, a brief description of the materials included in the packet, and other NSDUH field materials.

For persistent problem situations not resolved through FS and FI efforts or the individualized letters and packets, 7,217 "Call Me" letters were sent to the SDUs. Special care was taken to ensure that calls resulting from the letters were directed to the authorized RS or FS to set up an appointment so the FI could return and complete the screening.

Occasionally, controlled access problems required assistance beyond the RS level, so RDs—and sometimes even the National Field Director—became involved.

6.7 Refusal Conversion Procedures

More often than desired, potential respondents exercised their "right to refuse to participate." The following were in place to try to prevent refusal situations:

- All aspects of NSDUH were designed to exude professionalism and thus enhance the legitimacy of the project. All materials provided to the public were developed carefully. FIs were instructed to always behave professionally and courteously.
- The 2012 NSDUH Field Interviewer Manual gave specific instructions to FIs for introducing both themselves and the study. Additionally, an entire chapter discussed "Obtaining Participation" and listed the tools available to field staff along with tips for answering questions and overcoming objections.
- During new-to-project FI training, two sections of the guide covered details for contacting dwelling units and how to deal with reluctant respondents and difficult situations. During exercises and mock interviews, trainees were able to practice answering questions and using letters and handouts to obtain cooperation.
- During veteran FI training, time was spent reviewing various techniques for overcoming refusals. FIs participated in group discussions on completing screenings and interviews in different types of challenging refusal situations and sharing tips on avoiding and converting refusals in those situations. The exercises and ideas presented helped the FIs improve their skills and thus increase their confidence and ability to handle the many situations encountered in the field.

In refusal situations, staff followed these steps:

- Detailed notes describing the situation were recorded in a Refusal Report on the iPAQ. FIs selected the main reason for the refusal from the following categories:
  - Nothing in it for me
  - No time
  - Government/surveys too invasive
  - Gatekeeper/parent/household member won't allow participation
  - Confidentiality or survey legitimacy concerns
  - House too messy/too ill
  - Welfare/INS concerns
  - Need to discuss with FS
- After data transmission from the iPAQ to RTI, the category of refusal and any notes were then available to the FS on the CMS. The FI and FS could then discuss the situation, with the FS suggesting additional tactics if necessary.
- Once the refusal situation was discussed, a refusal conversion letter was sent (if appropriate). On the CMS, the FS selected a specific letter based on the stage of the

case (screening or interviewing), the category of the reason for the refusal, and, for interviewing, the person to be addressed (the actual respondent or the parent of a selected youth). The FS could also delete the request for the letter (in situations where a letter would not be helpful or could not be delivered) or release the letter for automatic production and mailing. During 2012, 37,589 refusal conversion letters were mailed.

- Supervisors were available to discuss the importance of participation with reluctant respondents.
- The FI returned to the SDU to try again with other tactics, except in the case of adamant or hostile refusals. If the FS determined a case was an adamant refusal based on discussions with the FI, the FS could choose to close the case without additional visits or transfer it to a different FI.

6.8 Problems Encountered

6.8.1 Size and Scope of the Project

By selecting areas throughout the entire country, many different types of situations arose that had to be resolved. With the large staff required by the size of the project, communication was vitally important, yet it was challenging to ensure that tips and suggestions were consistently conveyed to all staff.

6.8.2 Interviewing Staff Turnover

The continual turnover of interviewing staff meant there were not always enough FIs to adequately cover the assignments in all areas. Once replacement staff were in place, FSs needed to spend time mentoring these new FIs rather than focusing their attention on dealing with challenging cases. FSs spend a considerable amount of time dealing with staffing issues (recruiting, hiring, mentoring new employees, supervising new employees more intensely, implementing disciplinary actions with staff not meeting expectations, etc.), which ultimately taps into the amount of time they can allocate to managing the more difficult cases in their regions.

6.8.3 Refusals

Refusals at the screening and interviewing level have historically been a problem for NSDUH (as with all national-level household surveys). The introduction in 2002 of the $30 cash incentive for respondents completing the interview decreased the number of refusals and increased the number of interviews conducted in less than two visits to the household. However, FIs still had to deal with numerous issues in an effort to obtain cooperation:

- Some respondents refused because they felt they had already been inundated with market research and other survey requests.
- There was increased concern about providing personal information due to raised awareness of identity theft.

- Concerns about privacy and increased anti-government sentiment, including among immigrant populations, led to a larger portion of respondent refusals.

6.8.4 Typical Data Collection Concerns

As is common in any large field data collection effort, staff encountered problems such as respondent availability, dwelling unit access (controlled or otherwise restricted, particularly barriers such as fences, gates, or locked doors and/or "No Trespassing" signs), and high-crime neighborhoods. Additionally, the use of escorts to increase FI comfort levels in unsafe areas had an impact on respondent reactions.

6.8.5 Adverse Weather

Throughout the year, many regions across the nation experienced extreme weather conditions, which made travel difficult and created data collection challenges. To minimize the impact of these weather situations, field management implemented several strategies in order to complete field activities successfully once weather conditions improved. These strategies included encouraging FIs to work additional hours, utilizing traveling FIs and, in some areas, keeping screening cases open beyond the targeted deadline dates to improve screening response rates.

During Quarter 4, 2012, some segments in several Northeastern States were difficult to access or conduct field work in because of damage from Hurricane Sandy. Field costs, response rates, and fast start plans were all negatively impacted by the adverse conditions. Field management actively monitored conditions in these areas, and FIs resumed work as soon as circumstances allowed.

6.8.6 iPAQ

Using the iPAQ for electronic screening was a great use of technology, although the iPAQ had a few drawbacks:

- New staff unaccustomed to using computers needed time to build their confidence in using the iPAQ.
- Concentrating on the device meant less eye contact with the respondent, which in turn made it more challenging to establish good rapport.

6.8.7 CAI and iPAQ Patches

During data collection in 2012, one modification was made to the selection parameters for the Mental Health Surveillance Study by updating the CAI program loaded on FI laptops using a patch. Additionally, new transmission software for the FI laptop was released in Quarter 1, 2012. The new software allowed FIs to transmit data to RTI via analog phone line, Ethernet, or wireless internet connection. These updates did not affect the NSDUH CAI instrument.

No iPAQ patches were released during the 2012 data collection period.

Table 6.1 2012 NSDUH CAI Audit Trail Timing Data: Total Interview Time (Minutes) with FI Observation Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,465	45,696
Missing/Extreme Records		51	41	42	179	134	140
Summary Statistics (Minutes)							
Mean (μ)		61.8	61.8	61.7	63.0	63.7	63.4
Variance ($\sigma 2$)		252.1	241.0	247.5	356.0	366.9	363.2
Standard Deviation (σ)		15.9	15.5	15.7	18.9	19.2	19.1
Quartiles							
	Maximum	232.2	223.4	237.6	238.6	235.3	239.6
	Q3	70.3	69.9	69.9	72.3	72.7	72.6
	Median	59.4	59.6	59.5	59.6	60.1	59.9
	Q1	50.7	51.0	50.8	49.8	50.5	50.1
	Minimum	30.1	30.1	30.0	30.0	30.0	30.0
Range		202.2	193.3	207.6	208.6	205.3	209.6
Mode		52.8	49.1	58.4	51.3	49.6	55.4
Percentiles							
	99%	108.6	108.9	109.1	123.7	127.3	125.5
	95%	90.8	89.6	89.6	98.3	99.7	99.1
	90%	82.1	81.6	81.3	87.5	88.2	87.6
	10%	44.1	44.4	44.3	43.0	43.6	43.3
	5%	40.6	41.2	40.9	39.6	40.3	39.8
	1%	35.2	35.8	35.4	34.2	35.0	34.6
Extremes							
5 Highest	(Highest)	232.2	223.4	237.6	238.6	235.3	239.6
		227.9	212.7	233.9	238.3	223.9	238.5
		226.9	209.9	221.6	234.5	223.2	238.4
		218.4	205.2	214.6	226.0	222.1	236.2
		213.6	202.2	205.1	223.5	220.2	230.8
5 Lowest		30.6	30.5	30.3	30.2	30.1	30.1
		30.5	30.2	30.2	30.1	30.1	30.1
		30.3	30.1	30.2	30.1	30.1	30.0
		30.2	30.1	30.1	30.0	30.1	30.0
	(Lowest)	30.1	30.1	30.0	30.0	30.0	30.0

CAI = computer-assisted interviewing, FI = field interviewer.

Note: Time recording began at FIIDCON in the Introduction and stopped recording after FIEXIT in the FI Observation section of the Back-End Demographics Module.

Table 6.2 2012 NSDUH CAI Audit Trail Timing Data: Introduction and Core Demographics Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,465	45,696
Missing/Extreme Records		51	41	42	179	134	140
Summary Statistics (Minutes)							
Mean (μ)		4.8	4.9	4.7	4.9	5.0	4.7
Variance ($\sigma 2$)		7.6	8.3	8.8	9.2	10.3	11.5
Standard Deviation (σ)		2.8	2.9	3.0	3.0	3.2	3.4
Quartiles							
	Maximum	84.1	157.1	184.6	127.1	126.9	332.7
	Q3	5.8	5.8	5.6	5.6	5.7	5.4
	Median	4.3	4.4	4.2	4.2	4.3	4.1
	Q1	3.3	3.3	3.1	3.3	3.4	3.2
	Minimum	0.5	0.7	0.6	0.6	0.8	0.8
Range		83.7	156.4	184.0	126.5	126.1	331.9
Mode		3.4	4.0	4.0	4.1	3.5	4.0
Percentiles							
	99%	13.6	14.1	13.3	15.8	16.5	16.2
	95%	9.0	9.2	8.8	9.4	9.6	9.2
	90%	7.6	7.6	7.4	7.6	7.7	7.4
	10%	2.3	2.3	2.1	2.5	2.6	2.4
	5%	1.9	1.9	1.8	2.0	2.1	2.0
	1%	1.3	1.3	1.3	1.4	1.5	1.5
Extremes							
5 Highest	(Highest)	84.1	157.1	184.6	127.1	126.9	332.7
		78.7	77.2	103.8	71.2	104.7	90.3
		73.9	64.2	85.6	62.8	84.6	90.0
		71.1	50.6	72.2	62.1	80.0	68.4
		63.2	45.8	45.3	60.8	75.7	64.4
5 Lowest		0.7	0.8	0.8	0.8	0.8	0.9
		0.7	0.8	0.8	0.7	0.8	0.9
		0.7	0.7	0.8	0.7	0.8	0.9
		0.6	0.7	0.8	0.6	0.8	0.8
	(Lowest)	0.5	0.7	0.6	0.6	0.8	0.8

CAI = computer-assisted interviewing.

Note: Time recording began at FIIDCON in the Introduction and stopped recording after CALENDR3 in the Core Demographics Module.

Table 6.3 2012 NSDUH CAI Audit Trail Timing Data: Total ACASI

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,465	45,696
Missing/Extreme Records		51	41	42	179	134	140
Summary Statistics (Minutes)							
Mean (μ)		40.5	40.2	40.1	40.5	40.5	40.1
Variance ($\sigma 2$)		175.0	172.1	174.0	258.1	262.6	258.4
Standard Deviation (σ)		13.2	13.1	13.2	16.1	16.2	16.1
Quartiles							
	Maximum	172.4	182.6	187.4	223.8	197.3	223.8
	Q3	47.9	47.5	47.4	48.4	48.1	47.8
	Median	38.5	38.3	38.3	37.4	37.3	37.1
	Q1	31.1	30.9	30.8	29.2	29.2	28.9
	Minimum	8.0	8.9	7.5	8.9	5.2	8.1
Range		164.3	173.7	180.0	214.9	192.1	215.6
Mode		29.8	35.2	38.7	30.1	34.5	35.2
Percentiles							
	99%	78.9	79.8	79.3	91.8	93.3	92.4
	95%	65.0	63.7	64.0	70.9	71.2	70.4
	90%	58.0	56.9	57.0	61.5	61.7	60.9
	10%	25.6	25.5	25.4	23.5	23.6	23.2
	5%	22.8	22.5	22.6	20.6	20.7	20.4
	1%	18.1	18.1	17.9	16.1	16.0	15.9
Extremes							
5 Highest	(Highest)	172.4	182.6	187.4	223.8	197.3	223.8
		154.2	174.8	164.8	198.8	192.6	210.8
		145.6	148.6	143.0	178.0	188.2	196.1
		136.1	145.3	141.2	176.8	177.1	193.3
		131.3	142.0	137.4	175.8	175.2	167.1
5 Lowest		10.8	10.2	10.0	9.2	7.8	9.5
		10.5	9.4	9.9	9.0	7.6	9.5
		10.3	9.4	9.9	9.0	6.3	9.3
		9.7	9.3	7.5	9.0	5.9	8.9
	(Lowest)	8.0	8.9	7.5	8.9	5.2	8.1

ACASI = audio computer-assisted self-interviewing, CAI = computer-assisted interviewing.

Note: Time recording began at INTROACASI1 in the Beginning ACASI Module and stopped recording after ENDAUDIO in the Consumption of Alcohol Module.

Table 6.4 2012 NSDUH CAI Audit Trail Timing Data: Tutorial Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,465	45,696
Missing/Extreme Records		51	41	42	179	134	140
Summary Statistics (Minutes)							
Mean (μ)		6.1	6.1	6.1	5.8	5.8	5.8
Variance ($\sigma 2$)		3.5	3.3	3.1	5.1	5.2	4.9
Standard Deviation (σ)		1.9	1.8	1.7	2.3	2.3	2.2
Quartiles							
	Maximum	33.7	42.4	40.5	52.1	80.8	54.4
	Q3	7.2	7.1	7.0	6.9	6.8	6.7
	Median	6.0	5.9	6.0	5.5	5.5	5.4
	Q1	4.9	4.9	5.0	4.4	4.4	4.4
	Minimum	0.5	0.7	0.2	0.4	0.4	0.3
Range		33.2	41.7	40.3	51.7	80.4	54.1
Mode		6.2	5.8	6.2	4.8	5.2	4.8
Percentiles							
	99%	11.4	11.2	11.1	12.9	12.8	12.7
	95%	9.3	9.1	9.0	9.7	9.6	9.4
	90%	8.4	8.3	8.2	8.5	8.3	8.2
	10%	4.0	4.1	4.1	3.5	3.6	3.6
	5%	3.4	3.5	3.6	3.0	3.2	3.2
	1%	2.3	2.5	2.7	2.0	2.1	2.4
Extremes							
5 Highest	(Highest)	33.7	42.4	40.5	52.1	80.8	54.4
		27.8	37.8	30.1	47.9	74.0	51.8
		24.8	31.4	28.2	47.9	58.1	51.7
		24.5	24.8	26.4	45.8	53.7	48.3
		23.5	23.7	23.5	42.6	52.0	45.8
5 Lowest		0.8	0.8	1.2	0.5	0.6	0.5
		0.7	0.7	1.1	0.5	0.6	0.5
		0.7	0.7	1.0	0.5	0.4	0.4
		0.6	0.7	0.4	0.5	0.4	0.4
	(Lowest)	0.5	0.7	0.2	0.4	0.4	0.3

CAI = computer-assisted interviewing.

Note: Time recording began at INTROACASI in the Beginning ACASI Module and stopped recording after ANYQUES in the Tutorial Module.

Table 6.5 2012 NSDUH CAI Audit Trail Timing Data: Total Core Substances Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	41	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		11.9	12.0	12.0	12.6	12.8	12.7
Variance ($\sigma 2$)		29.5	29.4	28.3	37.6	39.5	37.6
Standard Deviation (σ)		5.4	5.4	5.3	6.1	6.3	6.1
Quartiles							
	Maximum	55.8	92.4	59.7	67.0	106.3	90.5
	Q3	15.0	15.0	15.1	15.5	15.6	15.4
	Median	10.9	11.0	11.0	11.2	11.4	11.4
	Q1	7.9	8.0	8.0	8.3	8.5	8.4
	Minimum	0.7	1.0	1.2	0.5	0.8	0.8
Range		55.1	91.5	58.5	66.5	105.5	89.7
Mode		9.8	9.2	7.6	10.1	8.9	9.3
Percentiles							
	99%	27.3	27.9	27.5	32.2	32.7	32.2
	95%	22.1	21.9	21.7	24.6	24.9	24.7
	90%	19.4	19.3	19.2	20.8	21.1	20.8
	10%	5.8	6.0	6.1	6.3	6.4	6.4
	5%	4.8	5.0	5.1	5.3	5.4	5.4
	1%	3.3	3.5	3.6	3.6	3.8	3.8
Extremes							
5 Highest	(Highest)	55.8	92.4	59.7	67.0	106.3	90.5
		52.9	64.5	56.8	64.6	87.0	78.7
		52.3	64.1	54.1	64.6	83.6	78.7
		49.6	62.4	52.5	63.3	74.9	77.6
		47.9	58.3	50.4	63.0	72.4	76.0
5 Lowest		1.2	1.2	1.4	0.5	1.1	1.2
		1.1	1.1	1.3	0.5	1.0	1.2
		1.0	1.1	1.3	0.5	1.0	1.2
		1.0	1.1	1.3	0.5	0.9	1.1
	(Lowest)	0.7	1.0	1.2	0.5	0.8	0.8

CAI = computer-assisted interviewing.

Note: Time recording began at LEADCIG in the Tobacco Module and stopped recording after SV13 in the Sedatives Module.

Table 6.6 2012 NSDUH CAI Audit Trail Timing Data: Tobacco Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	41	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		1.8	1.8	1.7	2.2	2.2	2.1
Variance ($\sigma 2$)		1.5	1.4	1.2	2.4	2.6	2.4
Standard Deviation (σ)		1.2	1.2	1.1	1.5	1.6	1.5
Quartiles							
	Maximum	29.5	22.1	17.3	24.6	39.7	36.7
	Q3	2.2	2.1	2.1	2.9	2.9	2.8
	Median	1.5	1.5	1.5	1.9	1.9	1.8
	Q1	1.0	1.0	1.0	1.1	1.1	1.0
	Minimum	0.2	0.1	0.2	0.0	0.1	0.1
Range		29.4	22.0	17.1	24.6	39.7	36.7
Mode		1.1	0.9	0.8	0.5	0.4	0.5
Percentiles							
	99%	6.4	6.3	5.9	7.3	7.4	7.3
	95%	4.2	4.1	3.8	5.0	5.0	4.9
	90%	3.2	3.1	2.9	4.1	4.1	4.0
	10%	0.7	0.7	0.7	0.6	0.6	0.6
	5%	0.6	0.6	0.6	0.4	0.4	0.4
	1%	0.5	0.5	0.5	0.3	0.3	0.3
Extremes							
5 Highest	(Highest)	29.5	22.1	17.3	24.6	39.7	36.7
		28.5	19.3	14.9	24.1	33.5	24.4
		17.9	13.8	14.0	23.3	31.9	24.1
		14.8	12.6	13.7	22.2	31.4	22.4
		14.0	12.5	12.6	19.8	30.9	21.9
5 Lowest		0.3	0.2	0.3	0.1	0.1	0.1
		0.2	0.2	0.3	0.1	0.1	0.1
		0.2	0.2	0.3	0.1	0.1	0.1
		0.2	0.2	0.2	0.1	0.1	0.1
	(Lowest)	0.2	0.1	0.2	0.0	0.1	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at LEADCIG and stopped recording after CG43 in the Tobacco Module.

Table 6.7 2012 NSDUH CAI Audit Trail Timing Data: Alcohol Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	41	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		1.7	1.6	1.6	2.4	2.5	2.4
Variance ($\sigma 2$)		1.4	1.4	1.3	2.1	2.2	2.2
Standard Deviation (σ)		1.2	1.2	1.1	1.5	1.5	1.5
Quartiles							
	Maximum	20.0	28.1	31.4	37.3	60.8	52.0
	Q3	2.3	2.2	2.2	3.1	3.1	3.1
	Median	1.4	1.4	1.4	2.2	2.2	2.2
	Q1	0.7	0.7	0.7	1.5	1.5	1.5
	Minimum	0.0	0.0	0.1	0.0	0.0	0.1
Range		20.0	28.0	31.4	37.3	60.7	52.0
Mode		2.3	2.3	2.2	1.9	2.3	2.2
Percentiles							
	99%	5.4	5.4	5.3	7.3	7.3	7.1
	95%	3.8	3.8	3.7	5.0	5.1	5.1
	90%	3.1	3.0	3.0	4.2	4.3	4.2
	10%	0.4	0.4	0.4	0.9	0.9	0.9
	5%	0.3	0.3	0.3	0.6	0.6	0.6
	1%	0.2	0.2	0.2	0.3	0.3	0.3
Extremes							
5 Highest	(Highest)	20.0	28.1	31.4	37.3	60.8	52.0
		14.7	20.1	22.4	28.4	33.7	33.1
		13.2	16.9	11.9	25.7	26.8	30.4
		11.0	15.8	11.0	22.3	26.6	29.4
		11.0	12.5	10.1	21.0	23.7	22.9
5 Lowest		0.1	0.1	0.1	0.1	0.1	0.1
		0.0	0.0	0.1	0.0	0.1	0.1
		0.0	0.0	0.1	0.0	0.1	0.1
		0.0	0.0	0.1	0.0	0.1	0.1
	(Lowest)	0.0	0.0	0.1	0.0	0.0	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at ALCINTR1 and stopped recording after ALCC30 in the Alcohol Module.

Table 6.8 2012 NSDUH CAI Audit Trail Timing Data: Marijuana Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,908	23,468	22,430	45,663	46,463	45,695
Missing/Extreme Records		52	42	43	181	136	141
Summary Statistics (Minutes)							
Mean (μ)		0.5	0.5	0.5	0.5	0.5	0.5
Variance ($\sigma 2$)		0.2	0.2	0.2	0.2	0.2	0.2
Standard Deviation (σ)		0.5	0.5	0.5	0.5	0.5	0.5
Quartiles							
	Maximum	9.5	12.6	6.9	20.3	21.5	18.8
	Q3	0.6	0.6	0.6	0.7	0.7	0.7
	Median	0.3	0.3	0.3	0.4	0.4	0.4
	Q1	0.2	0.2	0.2	0.2	0.2	0.2
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		9.5	12.6	6.9	20.3	21.5	18.8
Mode		0.1	0.1	0.1	0.1	0.1	0.1
Percentiles							
	99%	2.4	2.3	2.3	2.2	2.3	2.3
	95%	1.5	1.5	1.5	1.4	1.4	1.4
	90%	1.0	1.0	1.0	1.1	1.1	1.1
	10%	0.1	0.1	0.1	0.1	0.1	0.1
	5%	0.1	0.1	0.1	0.1	0.1	0.1
	1%	0.1	0.1	0.1	0.1	0.1	0.1
Extremes							
5 Highest	(Highest)	9.5	12.6	6.9	20.3	21.5	18.8
		6.7	11.8	6.5	8.7	12.6	11.9
		6.4	6.8	6.4	7.9	12.3	11.3
		5.6	5.7	6.3	7.6	10.6	8.8
		5.4	5.4	6.0	6.4	9.5	8.6
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at MRJINTRO and stopped recording after MJCC16 in the Marijuana Module.

Table 6.9 2012 NSDUH CAI Audit Trail Timing Data: Cocaine and Crack Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,907	23,468	22,429	45,664	46,461	45,692
Missing/Extreme Records		53	42	44	180	138	144
Summary Statistics (Minutes)							
Mean (μ)		0.2	0.2	0.2	0.2	0.2	0.2
Variance ($\sigma 2$)		0.0	0.0	0.0	0.1	0.1	0.1
Standard Deviation (σ)		0.2	0.2	0.2	0.3	0.3	0.3
Quartiles							
	Maximum	5.7	9.3	5.4	10.2	14.0	18.5
	Q3	0.2	0.2	0.2	0.3	0.3	0.3
	Median	0.1	0.1	0.1	0.1	0.1	0.1
	Q1	0.1	0.1	0.1	0.1	0.1	0.1
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		5.7	9.3	5.4	10.2	14.0	18.5
Mode		0.1	0.1	0.1	0.1	0.1	0.1
Percentiles							
	99%	0.9	0.8	0.7	1.6	1.5	1.5
	95%	0.4	0.4	0.4	0.8	0.7	0.7
	90%	0.3	0.3	0.3	0.5	0.5	0.5
	10%	0.1	0.1	0.1	0.1	0.1	0.1
	5%	0.1	0.1	0.1	0.1	0.1	0.1
	1%	0.0	0.0	0.0	0.0	0.0	0.0
Extremes							
5 Highest	(Highest)	5.7	9.3	5.4	10.2	14.0	18.5
		3.7	7.5	4.1	9.5	12.6	11.9
		3.3	4.4	3.8	7.4	11.3	11.6
		3.2	4.4	3.2	6.8	11.0	8.7
		2.9	3.9	3.0	6.4	11.0	8.2
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at COCINTRO in the Cocaine Module and stopped recording after CKCC16 in the Crack Module.

Table 6.10 2012 NSDUH CAI Audit Trail Timing Data: Heroin Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,900	23,456	22,417	45,627	46,428	45,666
Missing/Extreme Records		60	54	56	217	171	170
Summary Statistics (Minutes)							
Mean (μ)		0.1	0.1	0.1	0.1	0.1	0.1
Variance ($\sigma 2$)		0.0	0.0	0.0	0.0	0.0	0.0
Standard Deviation (σ)		0.1	0.1	0.1	0.1	0.1	0.2
Quartiles							
	Maximum	3.4	2.4	4.1	7.7	5.5	24.0
	Q3	0.1	0.1	0.1	0.1	0.1	0.1
	Median	0.1	0.1	0.1	0.1	0.1	0.1
	Q1	0.1	0.1	0.1	0.1	0.1	0.1
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		3.4	2.4	4.1	7.7	5.5	23.9
Mode		0.1	0.1	0.1	0.1	0.1	0.1
Percentiles							
	99%	0.3	0.3	0.3	0.5	0.5	0.5
	95%	0.2	0.2	0.2	0.2	0.2	0.2
	90%	0.2	0.2	0.2	0.2	0.2	0.2
	10%	0.1	0.1	0.1	0.1	0.1	0.0
	5%	0.0	0.0	0.0	0.0	0.0	0.0
	1%	0.0	0.0	0.0	0.0	0.0	0.0
Extremes							
5 Highest	(Highest)	3.4	2.4	4.1	7.7	5.5	24.0
		2.9	1.8	3.7	6.1	5.4	17.8
		2.7	1.6	2.2	5.8	3.7	10.1
		2.1	1.5	2.0	5.3	3.2	8.8
		2.1	1.4	1.8	5.0	2.9	7.5
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at HEINTRO and stopped recording after HECC16 in the Heroin Module.

Table 6.11 2012 NSDUH CAI Audit Trail Timing Data: Hallucinogens Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	41	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		0.9	0.9	0.9	0.8	0.8	0.8
Variance ($\sigma 2$)		0.4	0.4	0.3	0.5	0.5	0.6
Standard Deviation (σ)		0.6	0.6	0.6	0.7	0.7	0.8
Quartiles							
	Maximum	7.9	29.2	11.5	23.4	29.1	64.6
	Q3	1.2	1.2	1.2	1.0	1.0	1.0
	Median	0.8	0.7	0.7	0.6	0.6	0.6
	Q1	0.5	0.5	0.5	0.4	0.4	0.4
	Minimum	0.1	0.1	0.1	0.0	0.1	0.1
Range		7.8	29.1	11.5	23.4	29.1	64.6
Mode		0.4	0.5	0.5	0.4	0.4	0.4
Percentiles							
	99%	2.8	2.7	2.7	3.1	3.2	3.0
	95%	2.0	1.9	1.9	2.1	2.1	2.1
	90%	1.6	1.6	1.6	1.7	1.7	1.7
	10%	0.3	0.3	0.3	0.3	0.3	0.3
	5%	0.3	0.3	0.3	0.2	0.2	0.2
	1%	0.2	0.2	0.2	0.2	0.2	0.2
Extremes							
5 Highest	(Highest)	7.9	29.2	11.5	23.4	29.1	64.6
		7.7	11.3	9.2	15.9	28.3	31.2
		7.5	10.9	8.5	11.7	19.6	25.4
		7.0	10.1	7.9	11.7	18.7	25.1
		6.9	9.7	7.3	11.0	13.0	19.3
5 Lowest		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.0	0.1	0.1
	(Lowest)	0.1	0.1	0.1	0.0	0.1	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at HALINTRO and stopped recording after LSCC98 in the Hallucinogens Module.

Table 6.12 2012 NSDUH CAI Audit Trail Timing Data: Inhalants Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,469	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	41	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		1.4	1.4	1.4	1.1	1.1	1.1
Variance ($\sigma 2$)		0.9	1.0	0.9	0.8	1.1	0.8
Standard Deviation (σ)		1.0	1.0	0.9	0.9	1.0	0.9
Quartiles							
	Maximum	22.1	41.4	18.2	33.1	87.7	26.5
	Q3	1.8	1.8	1.8	1.3	1.3	1.3
	Median	1.2	1.1	1.1	0.9	0.9	0.9
	Q1	0.7	0.7	0.7	0.6	0.6	0.6
	Minimum	0.1	0.1	0.0	0.1	0.1	0.1
Range		22.1	41.3	18.1	33.0	87.6	26.4
Mode		0.7	0.8	0.6	0.6	0.6	0.6
Percentiles							
	99%	4.3	4.3	4.2	4.1	4.2	4.1
	95%	3.2	3.2	3.1	3.0	3.1	3.0
	90%	2.7	2.6	2.6	2.1	2.2	2.1
	10%	0.5	0.5	0.5	0.4	0.4	0.4
	5%	0.4	0.4	0.4	0.3	0.3	0.3
	1%	0.3	0.2	0.3	0.2	0.2	0.2
Extremes							
5 Highest	(Highest)	22.1	41.4	18.2	33.1	87.7	26.5
		18.5	21.7	10.9	19.8	29.7	18.6
		16.9	20.1	10.3	12.3	28.6	15.9
		12.2	18.8	10.0	12.2	27.2	15.7
		11.2	13.0	9.9	12.0	26.6	15.7
5 Lowest		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
	(Lowest)	0.1	0.1	0.0	0.1	0.1	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at INHINTRO and stopped recording after INCC16 in the Inhalants Module.

Table 6.13 2012 NSDUH CAI Audit Trail Timing Data: Total Pill Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,468	22,431	45,665	46,464	45,696
Missing/Extreme Records		51	42	42	179	135	140
Summary Statistics (Minutes)							
Mean (μ)		5.2	5.4	5.4	5.0	5.1	5.1
Variance (σ2)		7.5	7.6	7.3	8.1	8.5	8.2
Standard Deviation (σ)		2.7	2.8	2.7	2.8	2.9	2.9
Quartiles							
	Maximum	30.3	44.8	39.0	48.5	52.0	65.3
	Q3	6.7	6.8	6.9	6.3	6.4	6.3
	Median	4.7	4.9	4.9	4.4	4.5	4.5
	Q1	3.2	3.4	3.5	3.0	3.2	3.2
	Minimum	0.2	0.2	0.3	0.1	0.2	0.2
Range		30.1	44.6	38.8	48.4	51.8	65.1
Mode		4.9	3.3	4.2	3.3	3.7	3.6
Percentiles							
	99%	13.2	13.5	13.3	13.7	14.2	14.0
	95%	10.4	10.4	10.5	10.7	10.9	10.8
	90%	8.9	9.0	9.1	8.7	8.9	8.8
	10%	2.1	2.4	2.5	2.1	2.3	2.3
	5%	1.6	1.9	2.0	1.6	1.8	1.9
	1%	1.0	1.1	1.1	1.0	1.1	1.1
Extremes							
5 Highest	(Highest)	30.3	44.8	39.0	48.5	52.0	65.3
		27.5	33.3	28.1	43.9	49.6	45.3
		25.6	33.1	26.6	35.5	49.2	44.3
		22.1	30.8	25.5	33.0	45.0	41.5
		21.2	30.6	25.1	32.8	41.8	39.8
5 Lowest		0.3	0.3	0.4	0.2	0.3	0.3
		0.2	0.3	0.4	0.2	0.3	0.3
		0.2	0.3	0.3	0.2	0.3	0.3
		0.2	0.2	0.3	0.1	0.2	0.2
	(Lowest)	0.2	0.2	0.3	0.1	0.2	0.2

CAI = computer-assisted interviewing.

Note: Time recording began at INTROPR1 in the Pain Relievers Module and stopped recording after SV13 in the Sedatives Module.

Table 6.14 2012 NSDUH CAI Audit Trail Timing Data: Total Noncore Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,468	22,431	45,665	46,462	45,694
Missing/Extreme Records		51	42	42	179	137	142
Summary Statistics (Minutes)							
Mean (μ)		22.5	22.3	22.3	22.1	22.1	21.8
Variance ($\sigma 2$)		62.0	62.0	63.1	94.1	96.7	95.1
Standard Deviation (σ)		7.9	7.9	7.9	9.7	9.8	9.8
Quartiles							
	Maximum	129.2	128.1	132.0	191.3	114.6	156.9
	Q3	26.4	26.2	26.2	26.5	26.5	26.3
	Median	21.2	21.0	20.9	20.2	20.0	19.8
	Q1	17.1	16.9	16.8	15.4	15.4	15.1
	Minimum	3.9	2.6	2.3	3.2	3.0	2.6
Range		125.3	125.5	129.7	188.1	111.6	154.3
Mode		16.5	18.7	19.5	16.3	17.8	16.1
Percentiles							
	99%	47.0	47.4	47.7	54.1	55.5	54.4
	95%	37.0	36.6	37.0	40.1	40.5	39.9
	90%	32.6	32.1	32.2	34.4	34.4	34.1
	10%	14.0	13.9	13.8	12.1	12.1	11.9
	5%	12.5	12.4	12.2	10.5	10.5	10.3
	1%	9.7	9.6	9.6	8.0	7.9	7.8
Extremes							
5 Highest	(Highest)	129.2	128.1	132.0	191.3	114.6	156.9
		119.2	105.8	127.4	137.7	114.4	125.9
		118.7	95.8	88.7	126.2	113.1	122.6
		107.1	92.0	84.4	121.3	111.7	120.8
		91.8	87.2	83.0	114.5	106.0	113.1
5 Lowest		4.4	3.9	3.7	4.1	3.6	4.2
		4.4	3.4	3.2	3.8	3.6	4.1
		4.3	3.2	3.0	3.6	3.4	4.0
		4.0	3.0	3.0	3.5	3.2	3.8
	(Lowest)	3.9	2.6	2.3	3.2	3.0	2.6

CAI = computer-assisted interviewing.

Note: Time recording began at INTROSD in the Special Drugs Module and stopped recording after ENDAUDIO in the Consumption of Alcohol Module.

Table 6.15 2012 NSDUH CAI Audit Trail Timing Data: Special Drugs Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,468	22,431	45,665	46,462	45,694
Missing/Extreme Records		51	42	42	179	137	142
Summary Statistics (Minutes)							
Mean (μ)		1.7	1.7	1.7	1.6	1.6	1.6
Variance ($\sigma 2$)		0.6	0.6	0.5	0.8	0.9	0.9
Standard Deviation (σ)		0.8	0.7	0.7	0.9	0.9	0.9
Quartiles							
	Maximum	16.9	13.0	13.0	24.5	26.7	38.2
	Q3	2.1	2.1	2.1	1.9	2.0	1.9
	Median	1.6	1.6	1.6	1.4	1.4	1.4
	Q1	1.2	1.2	1.2	1.0	1.0	1.0
	Minimum	0.0	0.1	0.1	0.1	0.1	0.0
Range		16.8	12.9	12.9	24.4	26.7	38.1
Mode		1.6	1.2	1.4	1.1	1.1	1.1
Percentiles							
	99%	3.8	3.8	3.8	4.5	4.5	4.5
	95%	3.0	2.9	3.0	3.2	3.3	3.3
	90%	2.6	2.6	2.6	2.6	2.7	2.6
	10%	0.8	0.9	0.9	0.7	0.8	0.8
	5%	0.7	0.7	0.7	0.6	0.6	0.6
	1%	0.4	0.4	0.4	0.4	0.4	0.4
Extremes							
5 Highest	(Highest)	16.9	13.0	13.0	24.5	26.7	38.2
		10.5	11.9	8.9	23.6	23.5	28.5
		8.7	11.8	8.6	21.2	21.1	27.9
		8.5	10.8	7.4	16.8	20.8	24.6
		8.5	10.4	7.1	16.3	19.0	22.8
5 Lowest		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
		0.1	0.1	0.1	0.1	0.1	0.1
	(Lowest)	0.0	0.1	0.1	0.1	0.1	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at INTROSD and stopped recording after SD30 in the Special Drugs Module.

Table 6.16 2012 NSDUH CAI Audit Trail Timing Data: Risk/Availability Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		21,909	23,468	22,431	45,665	46,462	45,694
Missing/Extreme Records		51	42	42	179	137	142
Summary Statistics (Minutes)							
Mean (μ)		3.0	3.0	3.0	3.0	3.0	3.0
Variance ($\sigma 2$)		1.7	1.7	1.7	2.6	2.8	2.4
Standard Deviation (σ)		1.3	1.3	1.3	1.6	1.7	1.5
Quartiles							
	Maximum	48.1	25.0	25.6	67.2	80.7	55.8
	Q3	3.6	3.6	3.6	3.5	3.5	3.5
	Median	2.8	2.8	2.8	2.7	2.7	2.6
	Q1	2.1	2.1	2.1	2.0	2.0	2.0
	Minimum	0.5	0.1	0.3	0.1	0.2	0.1
Range		47.6	24.9	25.4	67.0	80.5	55.7
Mode		2.5	2.8	2.7	2.1	2.1	2.1
Percentiles							
	99%	7.4	7.4	7.6	8.6	8.7	8.5
	95%	5.3	5.3	5.3	5.8	5.9	5.8
	90%	4.5	4.5	4.6	4.7	4.7	4.7
	10%	1.7	1.7	1.7	1.6	1.6	1.6
	5%	1.5	1.5	1.5	1.4	1.4	1.4
	1%	1.1	1.1	1.2	1.1	1.1	1.1
Extremes							
5 Highest	(Highest)	48.1	25.0	25.6	67.2	80.7	55.8
		16.4	21.3	20.2	51.6	66.1	37.8
		15.5	20.7	19.8	42.4	50.9	28.9
		15.1	19.3	19.1	36.7	47.3	27.2
		14.9	16.3	18.4	31.9	42.0	25.4
5 Lowest		0.6	0.3	0.6	0.3	0.5	0.4
		0.6	0.3	0.5	0.3	0.4	0.4
		0.6	0.2	0.4	0.2	0.4	0.4
		0.6	0.2	0.4	0.2	0.2	0.3
	(Lowest)	0.5	0.1	0.3	0.1	0.2	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at RKQ1 and stopped recording after RK04d in the Risk/Availability Module.

Table 6.17 2012 NSDUH CAI Audit Trail Timing Data: Blunts Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,461	22,427	45,661	46,458	45,694
Missing/Extreme Records		51	49	46	183	141	142
Summary Statistics (Minutes)							
Mean (μ)		0.3	0.3	0.3	0.3	0.3	0.3
Variance ($\sigma 2$)		0.1	0.1	0.1	0.1	0.1	0.3
Standard Deviation (σ)		0.2	0.2	0.2	0.3	0.3	0.5
Quartiles							
	Maximum	3.7	6.9	5.5	20.0	26.3	97.2
	Q3	0.2	0.2	0.2	0.3	0.3	0.3
	Median	0.2	0.2	0.2	0.2	0.2	0.2
	Q1	0.2	0.2	0.2	0.2	0.2	0.2
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		3.7	6.9	5.5	20.0	26.3	97.2
Mode		0.2	0.2	0.2	0.1	0.1	0.2
Percentiles							
	99%	1.3	1.2	1.2	1.1	1.2	1.1
	95%	0.7	0.7	0.7	0.7	0.7	0.7
	90%	0.5	0.5	0.5	0.5	0.5	0.5
	10%	0.1	0.1	0.1	0.1	0.1	0.1
	5%	0.1	0.1	0.1	0.1	0.1	0.1
	1%	0.1	0.1	0.1	0.1	0.1	0.1
Extremes							
5 Highest	(Highest)	3.7	6.9	5.5	20.0	26.3	97.2
		3.5	3.9	4.0	19.7	14.4	12.1
		3.1	3.7	3.7	11.4	8.4	10.9
		3.0	3.6	3.5	5.4	6.1	10.9
		2.9	2.9	3.3	5.1	6.1	10.7
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at BL01 and stopped recording after BL07 in the Blunts Module.

Table 6.18 2012 NSDUH CAI Audit Trail Timing Data: Substance Dependence and Abuse Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		5,639	5,829	5,328	33,528	33,706	33,450
Missing/Extreme Records		16,321	17,681	17,145	12,316	12,893	12,386
Summary Statistics (Minutes)							
Mean (μ)		4.0	3.9	3.8	3.9	3.9	3.8
Variance ($\sigma 2$)		6.6	7.0	6.1	7.1	6.9	6.5
Standard Deviation (σ)		2.6	2.7	2.5	2.7	2.6	2.5
Quartiles							
	Maximum	23.2	28.2	25.2	60.1	39.7	46.0
	Q3	5.2	5.0	4.8	5.1	5.0	4.9
	Median	3.2	3.0	3.0	3.1	3.1	3.0
	Q1	2.2	2.1	2.1	2.0	2.0	2.0
	Minimum	0.2	0.2	0.2	0.1	0.1	0.2
Range		23.1	28.0	25.0	60.0	39.6	45.8
Mode		1.8	2.1	1.9	1.7	1.7	1.9
Percentiles							
	99%	12.3	13.0	12.4	13.0	13.1	12.4
	95%	9.1	8.9	8.6	8.9	8.7	8.5
	90%	7.5	7.4	7.0	7.3	7.1	7.0
	10%	1.6	1.5	1.5	1.5	1.5	1.5
	5%	1.3	1.3	1.2	1.3	1.3	1.2
	1%	0.7	0.6	0.7	0.9	0.9	0.9
Extremes							
5 Highest	(Highest)	23.2	28.2	25.2	60.1	39.7	46.0
		23.2	26.7	23.0	43.9	34.8	38.5
		22.3	23.4	22.4	36.4	34.5	34.1
		20.8	23.0	21.9	33.5	33.4	33.9
		18.6	22.9	20.7	33.4	32.6	33.7
5 Lowest		0.2	0.2	0.2	0.2	0.2	0.3
		0.2	0.2	0.2	0.2	0.2	0.3
		0.2	0.2	0.2	0.2	0.1	0.3
		0.2	0.2	0.2	0.2	0.1	0.2
	(Lowest)	0.2	0.2	0.2	0.1	0.1	0.2

CAI = computer-assisted interviewing.

Note: Time recording began at INTRODR and stopped recording after DRSV22 in the Substance Dependence and Abuse Module.

Table 6.19 2012 NSDUH CAI Audit Trail Timing Data: Market Information for Marijuana Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		3,033	3,250	3,058	9,157	9,253	9,451
Missing/Extreme Records		18,927	20,260	19,415	36,687	37,346	36,385
Summary Statistics (Minutes)							
Mean (μ)		1.5	1.5	1.5	1.5	1.5	1.5
Variance ($\sigma 2$)		0.4	0.4	0.4	0.5	0.5	0.6
Standard Deviation (σ)		0.7	0.6	0.6	0.7	0.7	0.7
Quartiles							
	Maximum	8.7	6.8	9.2	15.7	11.1	14.9
	Q3	1.8	1.8	1.8	1.8	1.9	1.8
	Median	1.4	1.4	1.4	1.4	1.4	1.4
	Q1	1.1	1.1	1.0	1.0	1.0	1.0
	Minimum	0.0	0.0	0.1	0.0	0.0	0.0
Range		8.7	6.8	9.1	15.7	11.1	14.9
Mode		1.5	1.1	1.2	1.1	1.5	1.1
Percentiles							
	99%	3.4	3.3	3.5	3.9	3.8	3.9
	95%	2.6	2.5	2.6	2.7	2.8	2.7
	90%	2.2	2.2	2.2	2.3	2.3	2.3
	10%	0.8	0.8	0.8	0.8	0.8	0.8
	5%	0.7	0.7	0.7	0.7	0.6	0.7
	1%	0.3	0.4	0.4	0.4	0.4	0.4
Extremes							
5 Highest	(Highest)	8.7	6.8	9.2	15.7	11.1	14.9
		8.5	6.4	6.5	15.2	10.6	13.8
		7.9	4.7	6.3	8.8	8.4	13.8
		7.3	4.6	5.3	8.6	8.2	12.4
		6.8	4.5	5.0	7.8	7.9	9.0
5 Lowest		0.1	0.1	0.1	0.0	0.1	0.1
		0.1	0.1	0.1	0.0	0.1	0.1
		0.0	0.0	0.1	0.0	0.1	0.1
		0.0	0.0	0.1	0.0	0.1	0.1
	(Lowest)	0.0	0.0	0.1	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at MJE01 and stopped recording after MJE70 in the Market Information for Marijuana Module.

Table 6.20 2012 NSDUH CAI Audit Trail Timing Data: Prior Substance Use Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		9,981	10,259	9,365	41,777	42,311	41,590
Missing/Extreme Records		11,979	13,251	13,108	4,067	4,288	4,246
Summary Statistics (Minutes)							
Mean (μ)		1.4	1.4	1.4	1.7	1.7	1.7
Variance ($\sigma 2$)		1.2	1.2	1.2	2.4	2.0	1.8
Standard Deviation (σ)		1.1	1.1	1.1	1.5	1.4	1.3
Quartiles							
	Maximum	20.5	11.3	9.7	157.7	60.1	28.6
	Q3	1.8	1.8	1.8	2.3	2.2	2.2
	Median	1.1	1.1	1.1	1.4	1.4	1.4
	Q1	0.6	0.6	0.6	0.8	0.8	0.8
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		20.5	11.3	9.6	157.7	60.0	28.6
Mode		0.8	0.6	0.7	1.0	0.5	0.8
Percentiles							
	99%	5.2	5.1	5.4	6.3	6.6	6.3
	95%	3.5	3.5	3.6	4.2	4.2	4.2
	90%	2.8	2.8	2.8	3.4	3.4	3.3
	10%	0.4	0.4	0.4	0.5	0.5	0.5
	5%	0.3	0.3	0.3	0.3	0.3	0.3
	1%	0.2	0.2	0.2	0.2	0.2	0.2
Extremes							
5 Highest	(Highest)	20.5	11.3	9.7	157.7	60.1	28.6
		16.5	11.0	9.6	38.8	38.9	22.8
		10.6	10.9	9.3	21.6	27.3	22.7
		10.3	10.1	9.3	21.2	24.4	19.5
		10.1	10.0	9.1	19.8	23.6	17.9
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at LU01 and stopped recording after LU39 in the Prior Substance Use Module.

Table 6.21 2012 NSDUH CAI Audit Trail Timing Data: Special Topics, Drug Treatment, and Health Care Sections

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,909	23,468	22,431	45,664	46,462	45,694
Missing/Extreme Records		51	42	42	180	137	142
Summary Statistics (Minutes)							
Mean (μ)		2.7	2.7	2.7	3.1	3.1	3.1
Variance ($\sigma 2$)		1.6	1.5	1.4	3.4	3.3	3.3
Standard Deviation (σ)		1.2	1.2	1.2	1.8	1.8	1.8
Quartiles							
	Maximum	32.1	41.5	21.7	86.3	39.1	53.2
	Q3	3.2	3.2	3.2	3.6	3.6	3.6
	Median	2.5	2.5	2.4	2.6	2.6	2.6
	Q1	2.0	1.9	1.9	2.0	2.0	2.0
	Minimum	0.1	0.1	0.0	0.1	0.0	0.1
Range		32.0	41.4	21.7	86.2	39.0	53.1
Mode		2.1	2.1	2.0	2.1	2.1	2.2
Percentiles							
	99%	7.2	7.1	7.0	9.7	9.7	9.7
	95%	4.9	4.9	4.8	6.4	6.4	6.4
	90%	4.2	4.1	4.1	5.1	5.1	5.1
	10%	1.6	1.6	1.6	1.6	1.6	1.6
	5%	1.4	1.4	1.4	1.4	1.4	1.4
	1%	1.1	1.1	1.1	1.1	1.1	1.1
Extremes							
5 Highest	(Highest)	32.1	41.5	21.7	86.3	39.1	53.2
		26.4	24.4	15.5	40.8	29.4	40.0
		20.0	19.6	14.1	37.7	29.3	37.5
		17.9	19.1	13.5	32.9	29.2	34.0
		17.7	16.2	13.4	31.4	28.9	32.5
5 Lowest		0.3	0.3	0.3	0.2	0.2	0.3
		0.2	0.2	0.3	0.2	0.2	0.2
		0.2	0.2	0.3	0.2	0.1	0.2
		0.2	0.2	0.1	0.1	0.1	0.1
	(Lowest)	0.1	0.1	0.0	0.1	0.0	0.1

CAI = computer-assisted interviewing.

Note: Time recording began at INTROSP in the Special Topics Module and stopped recording after CHK12M in the Health Care Module. The Market Information for Marijuana and Prior Substance Use Modules were embedded between Special Topics and Drug Treatment but were not included in these timing calculations.

Table 6.22 2011 NSDUH CAI Audit Trail Timing Data: Adult Mental Health Service Utilization Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		N/A	N/A	N/A	45,661	46,460	45,692
Missing/Extreme Records		N/A	N/A	N/A	183	139	144
Summary Statistics (Minutes)							
Mean (μ)		N/A	N/A	N/A	1.2	1.2	1.2
Variance ($\sigma 2$)		N/A	N/A	N/A	1.0	1.2	1.0
Standard Deviation (σ)		N/A	N/A	N/A	1.0	1.1	1.0
Quartiles							
	Maximum	N/A	N/A	N/A	46.2	62.9	37.1
	Q3	N/A	N/A	N/A	1.4	1.4	1.4
	Median	N/A	N/A	N/A	0.9	0.9	0.9
	Q1	N/A	N/A	N/A	0.7	0.7	0.7
	Minimum	N/A	N/A	N/A	0.0	0.0	0.0
Range		N/A	N/A	N/A	46.1	62.9	37.1
Mode		N/A	N/A	N/A	0.7	0.7	0.7
Percentiles							
	99%	N/A	N/A	N/A	5.1	5.3	5.2
	95%	N/A	N/A	N/A	3.0	3.1	3.0
	90%	N/A	N/A	N/A	2.3	2.4	2.3
	10%	N/A	N/A	N/A	0.5	0.5	0.5
	5%	N/A	N/A	N/A	0.4	0.4	0.4
	1%	N/A	N/A	N/A	0.2	0.2	0.2
Extremes							
5 Highest	(Highest)	N/A	N/A	N/A	46.2	62.9	37.1
		N/A	N/A	N/A	20.2	53.4	20.5
		N/A	N/A	N/A	18.4	25.9	19.3
		N/A	N/A	N/A	17.7	23.7	19.0
		N/A	N/A	N/A	17.4	19.1	15.7
5 Lowest		N/A	N/A	N/A	0.1	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
	(Lowest)	N/A	N/A	N/A	0.0	0.0	0.0

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at ADINTRO and stopped recording after ADMT30 in the Adult Mental Health Service Utilization Module.

Table 6.23 2012 NSDUH CAI Audit Trail Timing Data: Social Environment Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		N/A	N/A	N/A	45,660	46,459	45,692
Missing/Extreme Records		N/A	N/A	N/A	184	140	144
Summary Statistics (Minutes)							
Mean (μ)		N/A	N/A	N/A	1.5	1.5	1.5
Variance ($\sigma 2$)		N/A	N/A	N/A	0.8	0.7	0.7
Standard Deviation (σ)		N/A	N/A	N/A	0.9	0.8	0.8
Quartiles							
	Maximum	N/A	N/A	N/A	84.0	38.0	36.7
	Q3	N/A	N/A	N/A	1.7	1.7	1.7
	Median	N/A	N/A	N/A	1.3	1.3	1.3
	Q1	N/A	N/A	N/A	1.0	1.0	1.0
	Minimum	N/A	N/A	N/A	0.0	0.0	0.0
Range		N/A	N/A	N/A	83.9	38.0	36.7
Mode		N/A	N/A	N/A	1.2	1.2	1.0
Percentiles							
	99%	N/A	N/A	N/A	4.4	4.4	4.3
	95%	N/A	N/A	N/A	2.8	2.8	2.8
	90%	N/A	N/A	N/A	2.3	2.3	2.3
	10%	N/A	N/A	N/A	0.9	0.8	0.8
	5%	N/A	N/A	N/A	0.8	0.8	0.7
	1%	N/A	N/A	N/A	0.6	0.6	0.6
Extremes							
5 Highest	(Highest)	N/A	N/A	N/A	84.0	38.0	36.7
		N/A	N/A	N/A	34.4	23.4	30.8
		N/A	N/A	N/A	30.0	23.1	26.7
		N/A	N/A	N/A	25.3	21.9	24.4
		N/A	N/A	N/A	21.2	20.8	22.3
5 Lowest		N/A	N/A	N/A	0.1	0.0	0.0
		N/A	N/A	N/A	0.1	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
	(Lowest)	N/A	N/A	N/A	0.0	0.0	0.0

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at LEADSEN and stopped recording after SENREBE3 in the Social Environment Module.

Table 6.24 2012 NSDUH CAI Audit Trail Timing Data: Parenting Experiences Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		N/A	N/A	N/A	4,100	3,988	3,937
Missing/Extreme Records		N/A	N/A	N/A	41,744	42,611	41,899
Summary Statistics (Minutes)							
Mean (μ)		N/A	N/A	N/A	2.7	2.7	2.6
Variance ($\sigma 2$)		N/A	N/A	N/A	2.8	2.6	2.3
Standard Deviation (σ)		N/A	N/A	N/A	1.7	1.6	1.5
Quartiles							
	Maximum	N/A	N/A	N/A	56.2	26.5	21.3
	Q3	N/A	N/A	N/A	3.1	3.1	3.1
	Median	N/A	N/A	N/A	2.3	2.3	2.2
	Q1	N/A	N/A	N/A	1.7	1.8	1.7
	Minimum	N/A	N/A	N/A	0.0	0.1	0.1
Range		N/A	N/A	N/A	56.2	26.4	21.3
Mode		N/A	N/A	N/A	1.7	1.8	1.7
Percentiles							
	99%	N/A	N/A	N/A	8.2	8.7	8.0
	95%	N/A	N/A	N/A	5.5	5.7	5.5
	90%	N/A	N/A	N/A	4.4	4.5	4.3
	10%	N/A	N/A	N/A	1.5	1.4	1.4
	5%	N/A	N/A	N/A	1.3	1.3	1.3
	1%	N/A	N/A	N/A	1.0	1.0	0.9
Extremes							
5 Highest	(Highest)	N/A	N/A	N/A	56.2	26.5	21.3
		N/A	N/A	N/A	18.9	22.5	20.5
		N/A	N/A	N/A	17.3	17.1	20.1
		N/A	N/A	N/A	14.2	17.0	19.6
		N/A	N/A	N/A	12.5	16.7	17.2
5 Lowest		N/A	N/A	N/A	0.3	0.3	0.3
		N/A	N/A	N/A	0.3	0.3	0.3
		N/A	N/A	N/A	0.3	0.3	0.2
		N/A	N/A	N/A	0.2	0.2	0.2
	(Lowest)	N/A	N/A	N/A	0.0	0.1	0.1

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at LEADPAR and stopped recording after PE05d in the Parenting Experiences Module.

Table 6.25 2012 NSDUH CAI Audit Trail Timing Data: Youth Experiences Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		21,908	23,468	22,431	N/A	N/A	N/A
Missing/Extreme Records		52	42	42	N/A	N/A	N/A
Summary Statistics (Minutes)							
Mean (μ)		8.3	8.2	8.3	N/A	N/A	N/A
Variance ($\sigma 2$)		8.2	8.3	9.3	N/A	N/A	N/A
Standard Deviation (σ)		2.9	2.9	3.1	N/A	N/A	N/A
Quartiles							
	Maximum	49.1	44.8	107.8	N/A	N/A	N/A
	Q3	9.7	9.7	9.7	N/A	N/A	N/A
	Median	7.9	7.9	7.9	N/A	N/A	N/A
	Q1	6.4	6.3	6.3	N/A	N/A	N/A
	Minimum	0.2	0.2	0.2	N/A	N/A	N/A
Range		49.0	44.6	107.7	N/A	N/A	N/A
Mode		8.0	7.1	7.4	N/A	N/A	N/A
Percentiles							
	99%	17.3	17.3	17.6	N/A	N/A	N/A
	95%	13.2	13.2	13.4	N/A	N/A	N/A
	90%	11.8	11.6	11.8	N/A	N/A	N/A
	10%	5.2	5.2	5.1	N/A	N/A	N/A
	5%	4.6	4.5	4.5	N/A	N/A	N/A
	1%	3.2	3.2	3.2	N/A	N/A	N/A
Extremes							
5 Highest	(Highest)	49.1	44.8	107.8	N/A	N/A	N/A
		38.2	39.9	78.3	N/A	N/A	N/A
		37.9	38.8	51.0	N/A	N/A	N/A
		34.9	38.2	43.7	N/A	N/A	N/A
		34.2	37.2	38.4	N/A	N/A	N/A
5 Lowest		0.3	0.4	0.3	N/A	N/A	N/A
		0.3	0.4	0.3	N/A	N/A	N/A
		0.3	0.4	0.3	N/A	N/A	N/A
		0.2	0.3	0.3	N/A	N/A	N/A
	(Lowest)	0.2	0.2	0.2	N/A	N/A	N/A

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at LEADSEN and stopped recording after YEREBEL3 in the Youth Experiences Module.

Table 6.26 2012 NSDUH CAI Audit Trail Timing Data: Mental Health Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		N/A	N/A	N/A	45,657	46,456	45,693
Missing/Extreme Records		N/A	N/A	N/A	187	143	143
Summary Statistics (Minutes)							
Mean (μ)		N/A	N/A	N/A	3.2	3.2	3.2
Variance ($\sigma 2$)		N/A	N/A	N/A	4.5	4.3	4.4
Standard Deviation (σ)		N/A	N/A	N/A	2.1	2.1	2.1
Quartiles							
	Maximum	N/A	N/A	N/A	58.7	41.3	49.3
	Q3	N/A	N/A	N/A	4.1	4.2	4.1
	Median	N/A	N/A	N/A	2.9	2.9	2.9
	Q1	N/A	N/A	N/A	1.8	1.8	1.8
	Minimum	N/A	N/A	N/A	0.0	0.0	0.0
Range		N/A	N/A	N/A	58.7	41.3	49.3
Mode		N/A	N/A	N/A	2.8	2.4	2.5
Percentiles							
	99%	N/A	N/A	N/A	10.4	10.1	10.2
	95%	N/A	N/A	N/A	6.9	6.9	6.9
	90%	N/A	N/A	N/A	5.7	5.7	5.7
	10%	N/A	N/A	N/A	1.0	1.0	1.0
	5%	N/A	N/A	N/A	0.7	0.7	0.7
	1%	N/A	N/A	N/A	0.5	0.5	0.5
Extremes							
5 Highest	(Highest)	N/A	N/A	N/A	58.7	41.3	49.3
		N/A	N/A	N/A	44.3	35.0	40.9
		N/A	N/A	N/A	37.9	34.2	35.0
		N/A	N/A	N/A	34.2	30.4	34.6
		N/A	N/A	N/A	30.3	27.2	32.9
5 Lowest		N/A	N/A	N/A	0.1	0.1	0.0
		N/A	N/A	N/A	0.1	0.1	0.0
		N/A	N/A	N/A	0.1	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
	(Lowest)	N/A	N/A	N/A	0.0	0.0	0.0

CAI = computer-assisted interviewing, N/A = not applicable.

Note: The Mental Health Module included World Health Organization Disability Assessment Schedule (WHODAS) questions for the 2010, 2011, and 2012 NSDUHs.

Note: Time recording began at DIINTRO and stopped recording after SUI05 in the Mental Health Module.

Table 6.27 2012 NSDUH CAI Audit Trail Timing Data: Adult Depression Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		N/A	N/A	N/A	45,649	46,441	45,676
Missing/Extreme Records		N/A	N/A	N/A	195	158	160
Summary Statistics (Minutes)							
Mean (μ)		N/A	N/A	N/A	1.6	1.7	1.7
Variance ($\sigma 2$)		N/A	N/A	N/A	7.1	7.7	7.6
Standard Deviation (σ)		N/A	N/A	N/A	2.7	2.8	2.8
Quartiles							
	Maximum	N/A	N/A	N/A	34.0	78.4	38.9
	Q3	N/A	N/A	N/A	1.1	1.1	1.1
	Median	N/A	N/A	N/A	0.5	0.5	0.5
	Q1	N/A	N/A	N/A	0.3	0.3	0.3
	Minimum	N/A	N/A	N/A	0.0	0.0	0.0
Range		N/A	N/A	N/A	33.9	78.4	38.9
Mode		N/A	N/A	N/A	0.3	0.3	0.3
Percentiles							
	99%	N/A	N/A	N/A	12.1	12.4	12.2
	95%	N/A	N/A	N/A	7.7	7.8	7.9
	90%	N/A	N/A	N/A	5.7	5.8	5.8
	10%	N/A	N/A	N/A	0.2	0.2	0.2
	5%	N/A	N/A	N/A	0.2	0.2	0.2
	1%	N/A	N/A	N/A	0.1	0.1	0.1
Extremes							
5 Highest	(Highest)	N/A	N/A	N/A	34.0	78.4	38.9
		N/A	N/A	N/A	33.6	44.1	33.6
		N/A	N/A	N/A	26.9	37.1	30.5
		N/A	N/A	N/A	24.9	30.4	29.9
		N/A	N/A	N/A	23.8	29.3	28.1
5 Lowest		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
		N/A	N/A	N/A	0.0	0.0	0.0
	(Lowest)	N/A	N/A	N/A	0.0	0.0	0.0

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at ASC21 and stopped recording after AD86f in the Adult Depression Module.

Table 6.28 2012 NSDUH CAI Audit Trail Timing Data: Youth Mental Health Service Utilization Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,906	23,468	22,431	N/A	N/A	N/A
Missing/Extreme Records		54	42	42	N/A	N/A	N/A
Summary Statistics (Minutes)							
Mean (μ)		1.9	1.9	1.9	N/A	N/A	N/A
Variance ($\sigma 2$)		1.4	1.5	1.6	N/A	N/A	N/A
Standard Deviation (σ)		1.2	1.2	1.3	N/A	N/A	N/A
Quartiles							
	Maximum	24.7	27.6	23.0	N/A	N/A	N/A
	Q3	2.3	2.3	2.3	N/A	N/A	N/A
	Median	1.6	1.6	1.6	N/A	N/A	N/A
	Q1	1.1	1.1	1.1	N/A	N/A	N/A
	Minimum	0.0	0.0	0.0	N/A	N/A	N/A
Range		24.7	27.6	23.0	N/A	N/A	N/A
Mode		1.3	1.2	1.3	N/A	N/A	N/A
Percentiles							
	99%	6.1	6.3	6.5	N/A	N/A	N/A
	95%	4.0	4.1	4.2	N/A	N/A	N/A
	90%	3.3	3.3	3.4	N/A	N/A	N/A
	10%	0.8	0.8	0.8	N/A	N/A	N/A
	5%	0.6	0.6	0.6	N/A	N/A	N/A
	1%	0.3	0.3	0.3	N/A	N/A	N/A
Extremes							
5 Highest	(Highest)	24.7	27.6	23.0	N/A	N/A	N/A
		16.8	24.8	19.1	N/A	N/A	N/A
		16.2	18.5	18.7	N/A	N/A	N/A
		15.1	15.6	18.3	N/A	N/A	N/A
		14.6	15.5	18.1	N/A	N/A	N/A
5 Lowest		0.1	0.0	0.1	N/A	N/A	N/A
		0.0	0.0	0.1	N/A	N/A	N/A
		0.0	0.0	0.1	N/A	N/A	N/A
		0.0	0.0	0.1	N/A	N/A	N/A
	(Lowest)	0.0	0.0	0.0	N/A	N/A	N/A

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at INTROYSU and stopped recording after YSU36 in the Youth Mental Health Service Utilization Module.

Table 6.29 2012 NSDUH CAI Audit Trail Timing Data: Adolescent Depression Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,879	23,445	22,407	N/A	N/A	N/A
Missing/Extreme Records		81	65	66	N/A	N/A	N/A
Summary Statistics (Minutes)							
Mean (μ)		1.6	1.6	1.7	N/A	N/A	N/A
Variance ($\sigma 2$)		5.9	6.0	6.4	N/A	N/A	N/A
Standard Deviation (σ)		2.4	2.4	2.5	N/A	N/A	N/A
Quartiles							
	Maximum	22.0	27.4	27.8	N/A	N/A	N/A
	Q3	1.2	1.2	1.3	N/A	N/A	N/A
	Median	0.7	0.7	0.7	N/A	N/A	N/A
	Q1	0.4	0.4	0.4	N/A	N/A	N/A
	Minimum	0.0	0.0	0.0	N/A	N/A	N/A
Range		22.0	27.4	27.7	N/A	N/A	N/A
Mode		0.4	0.5	0.5	N/A	N/A	N/A
Percentiles							
	99%	10.7	10.7	10.8	N/A	N/A	N/A
	95%	7.5	7.6	7.7	N/A	N/A	N/A
	90%	5.5	5.7	5.8	N/A	N/A	N/A
	10%	0.2	0.2	0.3	N/A	N/A	N/A
	5%	0.2	0.2	0.2	N/A	N/A	N/A
	1%	0.0	0.0	0.0	N/A	N/A	N/A
Extremes							
5 Highest	(Highest)	22.0	27.4	27.8	N/A	N/A	N/A
		19.6	24.5	22.9	N/A	N/A	N/A
		19.5	20.6	21.1	N/A	N/A	N/A
		19.5	18.8	20.5	N/A	N/A	N/A
		17.5	18.6	20.2	N/A	N/A	N/A
5 Lowest		0.0	0.0	0.0	N/A	N/A	N/A
		0.0	0.0	0.0	N/A	N/A	N/A
		0.0	0.0	0.0	N/A	N/A	N/A
		0.0	0.0	0.0	N/A	N/A	N/A
	(Lowest)	0.0	0.0	0.0	N/A	N/A	N/A

CAI = computer-assisted interviewing, N/A = not applicable.

Note: Time recording began at YDS21 and stopped recording after YD86f in the Adolescent Depression Module.

Table 6.30 2012 NSDUH CAI Audit Trail Timing Data: Consumption of Alcohol Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		7,979	8,149	7,435	40,125	40,565	39,942
Missing/Extreme Records		13,981	15,361	15,038	5,719	6,034	5,894
Summary Statistics (Minutes)							
Mean (μ)		0.9	0.9	0.8	0.8	0.8	0.8
Variance ($\sigma 2$)		1.6	0.8	0.5	0.5	0.5	0.3
Standard Deviation (σ)		1.3	0.9	0.7	0.7	0.7	0.6
Quartiles							
	Maximum	93.7	38.6	10.1	49.3	75.3	45.0
	Q3	1.3	1.3	1.3	1.0	0.9	0.9
	Median	0.6	0.6	0.6	0.7	0.7	0.7
	Q1	0.3	0.3	0.3	0.5	0.5	0.5
	Minimum	0.0	0.0	0.0	0.0	0.0	0.0
Range		93.6	38.6	10.1	49.2	75.3	45.0
Mode		0.3	0.3	0.2	0.5	0.5	0.4
Percentiles							
	99%	3.2	3.2	3.1	2.7	2.6	2.6
	95%	2.2	2.2	2.2	1.7	1.7	1.7
	90%	1.8	1.8	1.8	1.4	1.4	1.3
	10%	0.2	0.2	0.2	0.3	0.3	0.3
	5%	0.2	0.2	0.2	0.3	0.3	0.3
	1%	0.1	0.1	0.1	0.2	0.2	0.2
Extremes							
5 Highest	(Highest)	93.7	38.6	10.1	49.3	75.3	45.0
		6.4	31.3	8.5	45.0	50.6	22.1
		5.5	6.6	8.4	33.6	20.3	12.6
		5.4	6.3	7.3	25.5	17.5	12.3
		4.9	5.9	6.5	24.2	17.0	11.8
5 Lowest		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
		0.0	0.0	0.0	0.0	0.0	0.0
	(Lowest)	0.0	0.0	0.0	0.0	0.0	0.0

CAI = computer-assisted interviewing.

Note: Time recording began at CAINTR and stopped recording after ENDAUDIO in the Consumption of Alcohol Module.

Table 6.31 2012 NSDUH CAI Audit Trail Timing Data: Total FI-Administered Back-End Demographics Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		21,908	23,466	22,430	45,656	46,450	45,693
Missing/Extreme Records		52	44	43	188	149	143
Summary Statistics (Minutes)							
Mean (μ)		11.3	11.4	11.4	12.6	12.7	12.7
Variance ($\sigma 2$)		26.7	29.2	24.4	27.9	27.4	24.9
Standard Deviation (σ)		5.2	5.4	4.9	5.3	5.2	5.0
Quartiles							
	Maximum	171.2	156.0	145.5	198.9	146.6	167.4
	Q3	13.3	13.3	13.3	14.7	14.8	14.8
	Median	10.5	10.6	10.6	11.9	12.0	12.0
	Q1	8.2	8.3	8.4	9.5	9.5	9.6
	Minimum	1.2	1.8	1.8	1.7	0.2	0.2
Range		170.0	154.2	143.7	197.3	146.4	167.2
Mode		9.7	9.8	9.3	10.2	11.0	10.7
Percentiles							
	99%	29.0	29.5	27.7	31.0	30.0	29.5
	95%	19.5	19.7	19.3	21.0	21.0	20.9
	90%	16.7	16.9	16.7	18.0	18.1	18.2
	10%	6.3	6.5	6.6	7.5	7.5	7.7
	5%	5.4	5.5	5.5	6.4	6.3	6.5
	1%	3.8	3.9	3.8	4.5	4.5	4.6
Extremes							
5 Highest	(Highest)	171.2	156.0	145.5	198.9	146.6	167.4
		114.4	153.6	111.7	122.4	125.5	97.8
		89.3	147.2	101.2	107.3	103.7	79.0
		88.2	122.8	97.6	84.0	97.6	78.4
		84.4	104.9	82.2	82.8	91.2	75.5
5 Lowest		2.0	2.2	2.2	2.0	2.0	1.9
		1.6	2.1	2.0	1.9	1.9	1.6
		1.4	2.0	2.0	1.9	1.7	1.6
		1.4	1.9	1.8	1.7	1.5	0.7
	(Lowest)	1.2	1.8	1.8	1.7	0.2	0.2

CAI = computer-assisted interviewing, FI = field interviewer.

Note: Time recording began at INTRODM2 and stopped recording after QI24 in the Back-End Demographics Module.

Table 6.32 2012 NSDUH CAI Audit Trail Timing Data: Back-End Demographics Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,908	23,466	22,430	45,656	46,450	45,693
Missing/Extreme Records		52	44	43	188	149	143
Summary Statistics (Minutes)							
Mean (μ)		4.9	5.0	5.1	7.2	7.3	7.4
Variance ($\sigma 2$)		7.3	8.4	8.7	11.7	12.0	12.2
Standard Deviation (σ)		2.7	2.9	2.9	3.4	3.5	3.5
Quartiles							
	Maximum	32.6	83.5	140.6	193.3	83.2	92.4
	Q3	6.1	6.2	6.3	8.8	8.9	9.0
	Median	4.3	4.3	4.4	6.9	7.0	7.1
	Q1	3.1	3.1	3.2	5.2	5.2	5.3
	Minimum	0.6	0.6	0.6	0.5	0.2	0.2
Range		32.0	82.9	140.0	192.9	83.0	92.2
Mode		3.6	3.8	3.8	6.8	7.3	6.5
Percentiles							
	99%	13.8	14.3	14.3	17.6	17.7	18.4
	95%	10.1	10.4	10.4	12.7	12.9	13.0
	90%	8.6	8.7	8.7	11.0	11.2	11.2
	10%	2.2	2.3	2.4	3.5	3.3	3.5
	5%	1.9	1.9	2.0	2.6	2.5	2.6
	1%	1.3	1.4	1.4	1.5	1.5	1.5
Extremes							
5 Highest	(Highest)	32.6	83.5	140.6	193.3	83.2	92.4
		28.3	70.8	59.3	76.8	78.9	67.2
		28.2	41.7	55.5	69.3	72.2	66.8
		28.0	41.3	39.8	67.8	66.4	60.6
		27.2	38.6	38.1	50.4	56.5	56.4
5 Lowest		0.7	0.8	0.7	0.5	0.5	0.5
		0.7	0.7	0.7	0.5	0.4	0.5
		0.7	0.7	0.7	0.5	0.4	0.4
		0.6	0.7	0.7	0.5	0.4	0.2
	(Lowest)	0.6	0.6	0.6	0.5	0.2	0.2

CAI = computer-assisted interviewing.

Note: Time recording began at INTRODM2 and stopped recording after SUPPGR30 in the Back-End Demographics Module.

Table 6.33 2012 NSDUH CAI Audit Trail Timing Data: Income Section

Age Category		12-17			18+		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		21,906	23,466	22,428	45,652	46,450	45,690
Missing/Extreme Records		54	44	45	192	149	146
Summary Statistics (Minutes)							
Mean (μ)		3.9	4.0	3.9	3.6	3.6	3.6
Variance ($\sigma 2$)		12.5	13.6	8.9	11.0	10.0	7.4
Standard Deviation (σ)		3.5	3.7	3.0	3.3	3.2	2.7
Quartiles							
	Maximum	169.6	151.2	109.4	114.2	117.7	163.0
	Q3	4.3	4.3	4.3	3.9	4.0	4.0
	Median	3.4	3.5	3.5	3.1	3.1	3.1
	Q1	2.7	2.7	2.8	2.4	2.4	2.5
	Minimum	0.1	0.3	0.3	0.2	0.1	0.3
Range		169.5	150.9	109.1	114.0	117.6	162.7
Mode		3.1	3.3	3.3	3.1	3.2	2.7
Percentiles							
	99%	17.9	19.4	15.9	18.5	16.8	14.2
	95%	7.2	7.4	6.9	6.6	6.6	6.4
	90%	5.7	5.6	5.5	5.2	5.2	5.1
	10%	1.8	1.9	1.9	1.8	1.8	1.9
	5%	1.4	1.4	1.4	1.4	1.5	1.5
	1%	0.8	0.8	0.8	0.9	0.9	1.0
Extremes							
5 Highest	(Highest)	169.6	151.2	109.4	114.2	117.7	163.0
		111.8	149.1	98.3	98.3	93.4	76.1
		81.0	98.8	91.4	74.4	89.7	73.1
		79.8	89.8	79.7	73.1	86.4	72.8
		74.5	72.9	62.6	68.8	76.5	67.0
5 Lowest		0.3	0.4	0.4	0.3	0.3	0.3
		0.3	0.4	0.4	0.3	0.3	0.3
		0.3	0.3	0.4	0.3	0.3	0.3
		0.3	0.3	0.3	0.3	0.3	0.3
	(Lowest)	0.1	0.3	0.3	0.2	0.1	0.3

CAI = computer-assisted interviewing.

Note: Time recording began at INTROINC and stopped recording after QI24 in the Income section of the Back-End Demographics Module.

Table 6.34 2012 NSDUH CAI Audit Trail Timing Data: FI Observation Section

Age Category		12-17			18+		
Year of Interest		2010	2011	2012	2010	2011	2012
Sample Used in Analysis		21,899	23,465	22,424	45,642	46,443	45,679
Missing/Extreme Records		61	45	49	202	156	157
Summary Statistics (Minutes)							
Mean (μ)		1.0	1.0	1.0	1.0	1.0	1.1
Variance ($\sigma 2$)		8.8	7.5	6.8	9.0	9.5	8.7
Standard Deviation (σ)		3.0	2.7	2.6	3.0	3.1	2.9
Quartiles							
	Maximum	177.6	139.7	155.8	161.2	176.1	147.6
	Q3	1.0	1.0	1.0	1.0	1.0	1.0
	Median	0.5	0.5	0.5	0.5	0.5	0.5
	Q1	0.3	0.3	0.3	0.3	0.3	0.3
	Minimum	0.0	0.1	0.1	0.0	0.0	0.0
Range		177.6	139.7	155.8	161.2	176.1	147.6
Mode		0.2	0.2	0.2	0.2	0.2	0.2
Percentiles							
	99%	7.4	7.7	7.7	8.5	8.8	9.4
	95%	2.8	2.7	2.8	2.9	2.9	3.0
	90%	1.8	1.8	1.9	1.9	1.9	1.9
	10%	0.2	0.2	0.2	0.2	0.2	0.2
	5%	0.2	0.2	0.2	0.2	0.2	0.2
	1%	0.1	0.1	0.1	0.1	0.1	0.1
Extremes							
5 Highest	(Highest)	177.6	139.7	155.8	161.2	176.1	147.6
		176.6	94.2	99.6	157.5	139.8	138.0
		163.2	87.8	93.4	143.3	138.6	122.6
		95.5	83.3	69.5	135.0	127.1	119.1
		76.4	83.2	66.9	134.7	115.8	100.4
5 Lowest		0.0	0.1	0.1	0.1	0.1	0.1
		0.0	0.1	0.1	0.1	0.1	0.1
		0.0	0.1	0.1	0.0	0.1	0.0
		0.0	0.1	0.1	0.0	0.1	0.0
	(Lowest)	0.0	0.1	0.1	0.0	0.0	0.0

CAI = computer-assisted interviewing, FI = field interviewer.

Note: Time recording began at FIDBFINTR and stopped recording after FIEXIT in the FI Observation section of the Back-End Demographics Module.

Table 6.35 2012 NSDUH CAI Audit Trail Timing Data: Back-End Demographics Section among Persons Aged 15 or Older, by Employment Status

Employment Status		Employed			Not Employed		
Year of Interest		**2010**	**2011**	**2012**	**2010**	**2011**	**2012**
Sample Used in Analysis		33,626	33,829	33,962	23,264	24,581	23,029
Missing/Extreme Records		117	90	83	101	84	77
Summary Statistics (Minutes)							
Mean (μ)		8.2	8.4	8.4	5.2	5.2	5.2
Variance ($\sigma 2$)		10.1	10.3	10.1	7.6	7.9	8.4
Standard Deviation (σ)		3.2	3.2	3.2	2.8	2.8	2.9
Quartiles							
	Maximum	193.3	83.2	67.2	76.8	48.7	92.4
	Q3	9.5	9.7	9.7	6.5	6.5	6.5
	Median	7.7	7.9	7.9	4.7	4.7	4.7
	Q1	6.3	6.4	6.5	3.3	3.3	3.4
	Minimum	1.0	1.2	0.2	0.5	0.2	0.2
Range		192.3	82.0	67.0	76.3	48.5	92.2
Mode		7.0	7.3	7.2	4.3	4.5	4.1
Percentiles							
	99%	18.4	18.6	19.3	14.0	14.4	14.6
	95%	13.5	13.6	13.8	10.0	10.2	10.1
	90%	11.7	11.9	12.0	8.5	8.6	8.5
	10%	5.2	5.3	5.4	2.4	2.4	2.4
	5%	4.7	4.7	4.8	1.9	1.9	1.9
	1%	3.7	3.7	3.8	1.2	1.2	1.3
Extremes							
5 Highest	(Highest)	193.3	83.2	67.2	76.8	48.7	92.4
		69.3	78.9	66.8	38.4	43.6	60.6
		67.8	72.2	56.4	37.6	42.1	55.5
		50.4	70.8	54.8	35.1	39.9	44.2
		49.1	66.4	52.3	34.7	39.5	43.9
5 Lowest		1.7	1.9	1.9	0.5	0.5	0.6
		1.4	1.8	1.8	0.5	0.4	0.5
		1.4	1.7	1.5	0.5	0.4	0.5
		1.2	1.7	1.3	0.5	0.4	0.4
	(Lowest)	1.0	1.2	0.2	0.5	0.2	0.2

CAI = computer-assisted interviewing.

Note: Time recording began at INTRODM2 and stopped recording after SUPPGR30 in the Back-End Demographics Module.

7. Data Collection Results

7.1 Overview

By following the data collection procedures already discussed, 214,274 dwelling units (DUs) were selected. As shown in Table 7.1, 178,586 units were identified as eligible during the screening process; that is, the units were not vacant or only occupied by active-duty military personnel, or other similar circumstances. From this number of eligible cases, 153,873 were then screened successfully. The selection procedure in the iPAQ yielded 87,656 eligible household members. From this number, a total of 68,309 interviews were then completed.

7.2 Screening Response Rates

The screening response rate is the total number of completed screenings divided by the total eligible DUs. The eligible DUs are computed by the sample dwelling units (SDUs) minus those SDUs not eligible to be included in the National Survey on Drug Use and Health (NSDUH). Ineligibles include vacant, not primary residence, not a DU, group quarters unit (GQU) listed as housing unit (HU), HU listed as GQU, only military, other ineligibles, and those SDUs where the residents will live there less than half of the quarter.

As a brief summary, Table 7.1 lists the sample totals and the national screening and interviewing response rates for the 2010, 2011, and 2012 surveys. Then, Tables 7.2 through 7.15 present the screening response rates for the 2012 sample nationwide. The final national screening response rates for the 2012 NSDUH were 86.16 percent (unweighted) and 86.07 percent (weighted).

Tables 7.2 and 7.3 show the national totals for ineligible and eligible cases, as broken down by population density and screening result code. Tables 7.4 and 7.5 display the national totals by complete and incomplete screening result code and population density. The next sets of tables list results for each State, broken down by population density (7.6 and 7.7), eligibility rate (7.8 and 7.9), completion rate (7.10 and 7.11), and nonresponse rate (7.12 and 7.13). Tables 7.14 and 7.15 show the reasons given for screening refusals for the national totals and then, in alphabetical order, for each State. Both unweighted and weighted tables are presented together for the Nation and each State.

7.3 Interview Response Rates

The interviewing response rate is the number of completed interviews divided by the total number of eligible respondents chosen through screening. If there are any ineligible respondents (e.g., adults in the military and youths younger than 12), these are subtracted from the total. The national rates for 2010, 2011, and 2012 are shown in Table 7.1.

Tables 7.16 through 7.29 present the interview response rates for the national sample. The final national interviewing response rates were 77.93 percent (unweighted) and 73.04 percent (weighted).

Tables 7.16 and 7.17 present the national unweighted and weighted interview response rates by smaller age groups for both 2011 and 2012. Tables 7.20 and 7.21 present the unweighted and weighted interview response rates for each State by age group. Both tables are presented on the same page for each State. Table 7.21a displays the national weighted interview response rates by age group and race/ethnicity. Tables 7.22 and 7.23 show national and State results of incomplete interviews by age, while Table 7.23a presents the national weighted results of incomplete interviews by both age and race/ethnicity. Tables 7.24 and 7.25 contain interview refusal reasons by age group for the Nation and for each State. Table 7.25a shows the weighted interview refusal reasons by age group and race/ethnicity for the Nation.

The remaining interview result tables are presented in pairs with the first table providing the unweighted percentages and the second table providing the weighted percentages. Tables 7.18 and 7.19 show the interview response rates by age group and gender. More detailed information by gender and smaller age groups is shown in Tables 7.26 and 7.27. Tables 7.28 and 7.29 present a summary of the interview response rates broken down by several factors including race/ethnicity, type of county, geographic region, and gender.

7.4 Spanish Interviews

The percentages of completed interviews that were conducted in Spanish are shown by State in Table 7.30 (unweighted) and Table 7.31 (weighted). Spanish interviewing percentages also were analyzed by age and county type in Table 7.32 (unweighted) and Table 7.33 (weighted). Tables 7.34 and 7.35 present the number of English- and Spanish-version interviews conducted by region and by population density, respectively.

7.5 Interviewer Assessment of the Interview

As part of each NSDUH interview, field interviewers (FIs) were required to assess the respondent's level of cooperation, understanding, and privacy during the interview. One question asked whether respondents revealed to the FI answers entered during the audio computer-assisted self-interviewing (ACASI) section.

All of these data were captured in the FI Observation Questions at the end of the interview and are summarized in Tables 7.36 through 7.39. These tables present data based on the FI's assessment of the respondent's level of understanding of the interview, the respondent's cooperation during the interview, the level of privacy during the interview, and how often the respondent revealed answers in the ACASI section. Each of these tables is broken down by the respondent's age and race/ethnicity.

7.6 Number of Visits

FIs were required to make at least four callback visits to DUs when attempting to complete screening and interviewing; however, callbacks continued to be made as long as the field supervisor (FS) felt there was a chance that the screening or the interview could be completed in a cost-effective manner. In some cases, more than 10 visits were made to complete a screening or interview. Tables 7.40 and 7.41 present data on the number of visits required to complete screenings and interviews.

Table 7.1 Summary of NSDUH Results

	2010		2011		2012	
Eligible DUs	166,532		179,293		178,586	
Complete Screenings	147,010		156,048		153,873	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Screening Response Rate	88.28	88.42	87.04	86.98	86.16	86.07
Selected Persons	84,997		88,536		87,656	
Completed Interviews	67,804		70,109		68,309	
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Interviewing Response Rate	79.77	74.57	79.19	74.38	77.93	73.04
	Unweighted	Weighted	Unweighted	Weighted	Unweighted	Weighted
Overall Response Rate	70.42	65.94	68.92	64.69	67.14	62.87

DUs = dwelling units.

Table 7.2 2012 Screening Results, by Population Density (Unweighted Percentages)

Screening Result	1,000,000+		50,000-999,999		Non-CBSA		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample	89,060	100.00	103,831	100.00	21,383	100.00	214,274	100.00
Ineligible Cases	12,094	13.58	17,421	16.78	6,173	28.87	35,688	16.66
Eligible Cases	76,966	86.42	86,410	83.22	15,210	71.13	178,586	83.34
Ineligibles	12,094	100.00	17,421	100.00	6,173	100.00	35,688	100.00
10 - Vacant	7,265	60.07	9,257	53.14	2,735	44.31	19,257	53.96
13 - Not Primary Residence	1,869	15.45	4,219	24.22	2,432	39.40	8,520	23.87
18 - Not a Dwelling Unit	872	7.21	1,162	6.67	462	7.48	2,496	6.99
22 - All Military Personnel	92	0.76	254	1.46	6	0.10	352	0.99
Other, Ineligible	1,996	16.50	2,529	14.52	538	8.72	5,063	14.19
Eligible Cases	76,966	100.00	86,410	100.00	15,210	100.00	178,586	100.00
Screening Complete	62,793	81.59	77,103	89.23	13,977	91.89	153,873	86.16
30 - No One Selected	37,161	48.28	46,674	54.01	9,156	60.20	92,991	52.07
31 - One Selected	13,801	17.93	16,849	19.50	2,805	18.44	33,455	18.73
32 - Two Selected	11,831	15.37	13,580	15.72	2,016	13.25	27,427	15.36
Screening Not Complete	14,173	18.41	9,307	10.77	1,233	8.11	24,713	13.84
11 - No One Home	1,923	2.50	948	1.10	158	1.04	3,029	1.70
12 - Respondent Unavailable	316	0.41	123	0.14	18	0.12	457	0.26
14 - Physically/Mentally Incompetent	292	0.38	263	0.30	42	0.28	597	0.33
15 - Language Barrier - Spanish	23	0.03	19	0.02	6	0.04	48	0.03
16 - Language Barrier - Other	619	0.80	124	0.14	5	0.03	748	0.42
17 - Refusal	8,834	11.48	7,083	8.20	890	5.85	16,807	9.41
21 - Other, Access Denied	1,777	2.31	527	0.61	55	0.36	2,359	1.32
24 - Other, Eligible	4	0.01	9	0.01	1	0.01	14	0.01
27 - Segment Not Accessible	0	0.00	0	0.00	0	0.00	0	0.00
33 - Screener Not Returned	45	0.06	38	0.04	7	0.05	90	0.05
39 - Fraudulent Case	339	0.44	173	0.20	51	0.34	563	0.32
44 - Electronic Screening Problem	1	0.00	0	0.00	0	0.00	1	0.00

CBSA = core-based statistical area.

Table 7.3 2012 Screening Results, by Population Density (Weighted Percentages)

Screening Result	1,000,000+		50,000-999,999		Non-CBSA		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample	89,060	100.00	103,831	100.00	21,383	100.00	214,274	100.00
Ineligible Cases	12,094	13.21	17,421	18.02	6,173	29.41	35,688	16.57
Eligible Cases	76,966	86.79	86,410	81.98	15,210	70.59	178,586	83.43
Ineligibles	12,094	100.00	17,421	100.00	6,173	100.00	35,688	100.00
10 - Vacant	7,265	57.00	9,257	49.52	2,735	42.89	19,257	51.50
13 - Not Primary Residence	1,869	19.33	4,219	29.63	2,432	42.74	8,520	27.46
18 - Not a Dwelling Unit	872	6.52	1,162	6.27	462	7.33	2,496	6.52
22 - All Military Personnel	92	0.89	254	1.29	6	0.10	352	0.97
Other, Ineligible	1,996	16.25	2,529	13.29	538	6.93	5,063	13.55
Eligible Cases	76,966	100.00	86,410	100.00	15,210	100.00	178,586	100.00
Screening Complete	62,793	82.68	77,103	89.23	13,977	91.62	153,873	86.07
30 - No One Selected	37,161	47.74	46,674	53.57	9,156	59.10	92,991	50.99
31 - One Selected	13,801	18.60	16,849	19.73	2,805	19.09	33,455	19.12
32 - Two Selected	11,831	16.33	13,580	15.92	2,016	13.43	27,427	15.96
Screening Not Complete	14,173	17.32	9,307	10.77	1,233	8.38	24,713	13.93
11 - No One Home	1,923	2.16	948	1.09	158	0.97	3,029	1.62
12 - Respondent Unavailable	316	0.37	123	0.16	18	0.07	457	0.26
14 - Physically/Mentally Incompetent	292	0.30	263	0.34	42	0.30	597	0.32
15 - Language Barrier - Spanish	23	0.03	19	0.03	6	0.06	48	0.03
16 - Language Barrier - Other	619	0.86	124	0.15	5	0.05	748	0.50
17 - Refusal	8,834	10.83	7,083	8.25	890	5.86	16,807	9.39
21 - Other, Access Denied	1,777	2.19	527	0.53	55	0.59	2,359	1.37
24 - Other, Eligible	4	0.00	9	0.01	1	0.01	14	0.01
27 - Segment Not Accessible	0	0.00	0	0.00	0	0.00	0	0.00
33 - Screener Not Returned	45	0.06	38	0.04	7	0.04	90	0.05
39 - Fraudulent Case	339	0.54	173	0.17	51	0.43	563	0.37
44 - Electronic Screening Problem	1	0.00	0	0.00	0	0.00	1	0.00

CBSA = core-based statistical area.

Table 7.4 2012 Screening Results, by Final Result and Population Density (Unweighted Percentages)

	1,000,000+		50,000-999,999		Non-CBSA		Total	
Screening Result	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Screening Complete	62,793	100.00	77,103	100.00	13,977	100.00	153,873	100.00
30 - No One Selected	37,161	59.18	46,674	60.53	9,156	65.51	92,991	60.43
31 - One Selected	13,801	21.98	16,849	21.85	2,805	20.07	33,455	21.74
32 - Two Selected	11,831	18.84	13,580	17.61	2,016	14.42	27,427	17.82
Screening Not Complete	14,173	100.00	9,307	100.00	1,233	100.00	24,713	100.00
11 - No One Home	1,923	13.57	948	10.19	158	12.81	3,029	12.26
12 - Respondent Unavailable	316	2.23	123	1.32	18	1.46	457	1.85
14 - Physically/Mentally Incompetent	292	2.06	263	2.83	42	3.41	597	2.42
15 - Language Barrier - Spanish	23	0.16	19	0.20	6	0.49	48	0.19
16 - Language Barrier - Other	619	4.37	124	1.33	5	0.41	748	3.03
17 - Refusal	8,834	62.33	7,083	76.10	890	72.18	16,807	68.01
21 - Other, Access Denied	1,777	12.54	527	5.66	55	4.46	2,359	9.55
24 - Other, Eligible	4	0.03	9	0.10	1	0.08	14	0.06
27 - Segment Not Accessible	0	0.00	0	0.00	0	0.00	0	0.00
33 - Screener Not Returned	45	0.32	38	0.41	7	0.57	90	0.36
39 - Fraudulent Case	339	2.39	173	1.86	51	4.14	563	2.28
44 - Electronic Screening Problem	1	0.01	0	0.00	0	0.00	1	0.00

CBSA = core-based statistical area.

Table 7.5 2012 Screening Results, by Final Result and Population Density (Weighted Percentages)

	1,000,000+		50,000-999,999		Non-CBSA		Total	
Screening Result	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Screening Complete	62,793	100.00	77,103	100.00	13,977	100.00	153,873	100.00
30 - No One Selected	37,161	57.75	46,674	60.04	9,156	64.50	92,991	59.24
31 - One Selected	13,801	22.50	16,849	22.12	2,805	20.83	33,455	22.21
32 - Two Selected	11,831	19.75	13,580	17.85	2,016	14.66	27,427	18.55
Screening Not Complete	14,173	100.00	9,307	100.00	1,233	100.00	24,713	100.00
11 - No One Home	1,923	12.44	948	10.11	158	11.58	3,029	11.63
12 - Respondent Unavailable	316	2.13	123	1.47	18	0.88	457	1.86
14 - Physically/Mentally Incompetent	292	1.72	263	3.19	42	3.58	597	2.28
15 - Language Barrier - Spanish	23	0.17	19	0.27	6	0.76	48	0.23
16 - Language Barrier - Other	619	4.95	124	1.41	5	0.56	748	3.60
17 - Refusal	8,834	62.49	7,083	76.56	890	69.92	16,807	67.44
21 - Other, Access Denied	1,777	12.63	527	4.89	55	7.03	2,359	9.84
24 - Other, Eligible	4	0.02	9	0.10	1	0.11	14	0.05
27 - Segment Not Accessible	0	0.00	0	0.00	0	0.00	0	0.00
33 - Screener Not Returned	45	0.35	38	0.39	7	0.51	90	0.37
39 - Fraudulent Case	339	3.09	173	1.62	51	5.08	563	2.68
44 - Electronic Screening Problem	1	0.01	0	0.00	0	0.00	1	0.01

CBSA = core-based statistical area.

Table 7.6 2012 Screening Results—Completion Rate, by State and Population Density (Unweighted Percentages)

	1,000,000+		50,000-999,999		Non-CBSA		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Total	62,793	81.59	77,103	89.23	13,977	91.89	153,873	86.16
Alabama	460	90.55	1,461	90.07	220	90.91	2,141	90.26
Alaska	0	0.00	1,262	88.50	380	85.78	1,642	87.85
Arizona	1,239	88.94	662	92.59	27	77.14	1,928	89.97
Arkansas	23	88.46	1,674	90.10	393	96.32	2,090	91.19
California	5,241	81.88	1,568	84.26	43	82.69	6,852	82.42
Colorado	951	80.94	1,047	87.91	203	95.31	2,201	85.34
Connecticut	707	82.40	1,400	83.48	0	0.00	2,107	83.12
Delaware	0	0.00	2,008	87.61	0	0.00	2,008	87.61
District of Columbia	3,327	81.07	0	0.00	0	0.00	3,327	81.07
Florida	5,042	82.74	3,371	87.60	103	91.15	8,516	84.69
Georgia	974	87.51	703	87.66	119	93.70	1,796	87.95
Hawaii	0	0.00	2,239	81.09	0	0.00	2,239	81.09
Idaho	0	0.00	1,640	93.88	181	94.27	1,821	93.91
Illinois	4,669	73.12	2,530	83.31	479	88.38	7,678	77.06
Indiana	527	90.86	1,270	91.10	124	91.18	1,921	91.04
Iowa	0	0.00	1,524	91.20	498	94.32	2,022	91.95
Kansas	577	89.18	1,124	89.92	276	91.69	1,977	89.95
Kentucky	816	90.97	864	93.10	522	89.69	2,202	91.48
Louisiana	402	88.74	1,490	93.18	85	93.41	1,977	92.25
Maine	0	0.00	1,736	89.16	849	93.19	2,585	90.45
Maryland	1,527	77.79	260	79.51	15	83.33	1,802	78.08
Massachusetts	1,289	81.53	919	85.73	0	0.00	2,208	83.23
Michigan	3,084	80.29	4,129	88.23	613	89.36	7,826	85.00
Minnesota	1,262	90.21	461	94.27	252	92.65	1,975	91.44
Mississippi	180	92.78	1,399	93.33	372	94.42	1,951	93.48

Table 7.6 2012 Screening Results—Completion Rate, by State and Population Density (Unweighted Percentages) (continued)

State	1,000,000+		50,000-999,999		Non-CBSA		Total	
	Count	%	Count	%	Count	%	Count	%
Missouri	1,177	91.17	666	89.28	345	92.74	2,188	90.83
Montana	0	0.00	1,538	91.71	877	94.00	2,415	92.53
Nebraska	0	0.00	1,480	93.08	538	91.97	2,018	92.78
Nevada	1,104	91.85	557	91.46	60	88.24	1,721	91.59
New Hampshire	0	0.00	1,906	87.15	285	89.06	2,191	87.40
New Jersey	1,718	86.12	217	93.53	0	0.00	1,935	86.89
New Mexico	0	0.00	1,756	92.03	133	92.36	1,889	92.06
New York	6,282	68.27	2,582	84.57	251	85.96	9,115	72.65
North Carolina	261	85.57	1,512	90.05	217	82.82	1,990	88.60
North Dakota	0	0.00	1,652	93.65	809	93.10	2,461	93.47
Ohio	4,284	86.88	4,292	90.78	447	96.54	9,023	89.14
Oklahoma	700	89.74	1,120	91.65	353	92.89	2,173	91.23
Oregon	900	89.02	1,021	90.19	98	91.59	2,019	89.73
Pennsylvania	4,007	76.82	4,203	87.98	243	92.40	8,453	82.42
Rhode Island	1,957	89.36	0	0.00	0	0.00	1,957	89.36
South Carolina	109	84.50	2,146	89.27	119	89.47	2,374	89.05
South Dakota	0	0.00	1,555	93.56	476	95.01	2,031	93.90
Tennessee	792	90.41	920	93.12	217	93.94	1,929	92.08
Texas	4,277	86.72	2,016	92.14	499	93.97	6,792	88.77
Utah	0	0.00	1,393	94.50	81	96.43	1,474	94.61
Vermont	0	0.00	1,698	88.62	619	85.85	2,317	87.86
Virginia	1,324	87.74	585	88.91	118	93.65	2,027	88.40
Washington	1,031	88.88	961	91.18	86	93.48	2,078	90.11
West Virginia	56	91.80	1,900	88.58	443	94.46	2,399	89.68
Wisconsin	517	91.67	1,104	92.93	269	93.08	1,890	92.60
Wyoming	0	0.00	1,582	91.13	640	92.89	2,222	91.63

CBSA = core-based statistical area.

Table 7.7 2012 Screening Results—Completion Rate, by State and Population Density (Weighted Percentages)

	1,000,000+		50,000-999,999		Non-CBSA		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Total	62,793	82.68	77,103	89.23	13,977	91.62	153,873	86.07
Alabama	460	90.93	1,461	89.99	220	91.07	2,141	90.30
Alaska	0	0.00	1,262	88.49	380	85.71	1,642	87.82
Arizona	1,239	89.36	662	92.39	27	77.23	1,928	90.16
Arkansas	23	88.46	1,674	89.91	393	96.37	2,090	90.92
California	5,241	81.84	1,568	84.19	43	82.74	6,852	82.37
Colorado	951	81.46	1,047	86.28	203	95.94	2,201	85.23
Connecticut	707	81.94	1,400	83.20	0	0.00	2,107	82.76
Delaware	0	0.00	2,008	87.57	0	0.00	2,008	87.57
District of Columbia	3,327	80.90	0	0.00	0	0.00	3,327	80.90
Florida	5,042	82.62	3,371	87.52	103	91.27	8,516	84.67
Georgia	974	87.39	703	87.83	119	93.74	1,796	87.94
Hawaii	0	0.00	2,239	80.80	0	0.00	2,239	80.80
Idaho	0	0.00	1,640	93.92	181	93.95	1,821	93.92
Illinois	4,669	73.09	2,530	83.18	479	88.49	7,678	77.04
Indiana	527	90.59	1,270	91.14	124	91.36	1,921	91.01
Iowa	0	0.00	1,524	90.96	498	94.36	2,022	91.72
Kansas	577	89.26	1,124	90.00	276	91.43	1,977	89.98
Kentucky	816	91.13	864	93.11	522	89.32	2,202	91.46
Louisiana	402	88.51	1,490	93.29	85	93.44	1,977	92.28
Maine	0	0.00	1,736	89.25	849	93.20	2,585	90.56
Maryland	1,527	77.82	260	79.66	15	83.33	1,802	78.13
Massachusetts	1,289	81.47	919	85.78	0	0.00	2,208	83.22
Michigan	3,084	80.41	4,129	88.21	613	89.10	7,826	85.05
Minnesota	1,262	90.43	461	94.40	252	92.80	1,975	91.57
Mississippi	180	93.02	1,399	93.34	372	94.33	1,951	93.50

Table 7.7 2012 Screening Results—Completion Rate, by State and Population Density (Weighted Percentages) (continued)

	1,000,000+		50,000-999,999		Non-CBSA		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**	**Count**	**%**
Missouri	1,177	91.24	666	89.42	345	92.65	2,188	90.88
Montana	0	0.00	1,538	91.82	877	94.06	2,415	92.62
Nebraska	0	0.00	1,480	93.05	538	91.88	2,018	92.74
Nevada	1,104	92.04	557	91.50	60	89.09	1,721	91.75
New Hampshire	0	0.00	1,906	87.18	285	88.86	2,191	87.40
New Jersey	1,718	86.09	217	93.45	0	0.00	1,935	86.87
New Mexico	0	0.00	1,756	92.17	133	92.83	1,889	92.22
New York	6,282	67.28	2,582	84.54	251	85.24	9,115	71.89
North Carolina	261	85.64	1,512	90.12	217	81.51	1,990	88.48
North Dakota	0	0.00	1,652	93.55	809	93.17	2,461	93.42
Ohio	4,284	86.85	4,292	90.82	447	96.58	9,023	89.14
Oklahoma	700	89.93	1,120	91.50	353	92.97	2,173	91.22
Oregon	900	88.92	1,021	89.93	98	91.77	2,019	89.57
Pennsylvania	4,007	76.38	4,203	87.90	243	91.99	8,453	82.09
Rhode Island	1,957	89.37	0	0.00	0	0.00	1,957	89.37
South Carolina	109	84.62	2,146	89.16	119	89.41	2,374	88.97
South Dakota	0	0.00	1,555	93.55	476	95.08	2,031	93.92
Tennessee	792	90.27	920	92.90	217	94.06	1,929	91.91
Texas	4,277	86.48	2,016	92.16	499	93.96	6,792	88.52
Utah	0	0.00	1,393	94.58	81	96.46	1,474	94.67
Vermont	0	0.00	1,698	88.47	619	85.95	2,317	87.81
Virginia	1,324	87.89	585	88.77	118	93.63	2,027	88.47
Washington	1,031	89.20	961	90.80	86	93.37	2,078	90.10
West Virginia	56	92.05	1,900	88.14	443	94.75	2,399	89.39
Wisconsin	517	91.34	1,104	92.71	269	92.98	1,890	92.37
Wyoming	0	0.00	1,582	91.20	640	93.05	2,222	91.72

CBSA = core-based statistical area.

Table 7.8 2012 Screening Results—Eligibility Rate, by State (Unweighted Percentages)

State	SDUs	Eligible DUs	% Eligible DUs
Total	214,274	178,586	83.34
Alabama	3,012	2,372	78.75
Alaska	2,424	1,869	77.10
Arizona	2,771	2,143	77.34
Arkansas	2,776	2,292	82.56
California	9,489	8,314	87.62
Colorado	3,071	2,579	83.98
Connecticut	2,855	2,535	88.79
Delaware	2,847	2,292	80.51
District of Columbia	5,055	4,104	81.19
Florida	12,768	10,055	78.75
Georgia	2,365	2,042	86.34
Hawaii	3,212	2,761	85.96
Idaho	2,300	1,939	84.30
Illinois	11,385	9,964	87.52
Indiana	2,491	2,110	84.70
Iowa	2,529	2,199	86.95
Kansas	2,598	2,198	84.60
Kentucky	2,852	2,407	84.40
Louisiana	2,741	2,143	78.18
Maine	3,866	2,858	73.93
Maryland	2,680	2,308	86.12
Massachusetts	3,064	2,653	86.59
Michigan	11,441	9,207	80.47
Minnesota	2,483	2,160	86.99
Mississippi	2,553	2,087	81.75

Table 7.8 2012 Screening Results—Eligibility Rate, by State (Unweighted Percentages) (continued)

State	SDUs	Eligible DUs	% Eligible DUs
Missouri	2,879	2,409	83.67
Montana	3,295	2,610	79.21
Nebraska	2,556	2,175	85.09
Nevada	2,354	1,879	79.82
New Hampshire	2,990	2,507	83.85
New Jersey	2,622	2,227	84.94
New Mexico	2,771	2,052	74.05
New York	14,547	12,547	86.25
North Carolina	2,848	2,246	78.86
North Dakota	3,374	2,633	78.04
Ohio	11,722	10,122	86.35
Oklahoma	2,960	2,382	80.47
Oregon	2,547	2,250	88.34
Pennsylvania	11,907	10,256	86.13
Rhode Island	2,620	2,190	83.59
South Carolina	3,306	2,666	80.64
South Dakota	2,636	2,163	82.06
Tennessee	2,532	2,095	82.74
Texas	9,048	7,651	84.56
Utah	1,793	1,558	86.89
Vermont	3,292	2,637	80.10
Virginia	2,576	2,293	89.01
Washington	2,700	2,306	85.41
West Virginia	3,222	2,675	83.02
Wisconsin	2,440	2,041	83.65
Wyoming	3,109	2,425	78.00

DUs = dwelling units, SDUs = sample dwelling units.

Table 7.9 2012 Screening Results—Eligibility Rate, by State (Weighted Percentages)

State	SDUs	Eligible DUs	% Eligible DUs
Total	214,274	178,586	83.43
Alabama	3,012	2,372	78.65
Alaska	2,424	1,869	76.83
Arizona	2,771	2,143	76.44
Arkansas	2,776	2,292	82.70
California	9,489	8,314	86.08
Colorado	3,071	2,579	83.17
Connecticut	2,855	2,535	88.73
Delaware	2,847	2,292	78.08
District of Columbia	5,055	4,104	80.88
Florida	12,768	10,055	75.81
Georgia	2,365	2,042	86.31
Hawaii	3,212	2,761	85.80
Idaho	2,300	1,939	84.78
Illinois	11,385	9,964	87.57
Indiana	2,491	2,110	84.55
Iowa	2,529	2,199	86.56
Kansas	2,598	2,198	84.94
Kentucky	2,852	2,407	84.44
Louisiana	2,741	2,143	77.93
Maine	3,866	2,858	73.00
Maryland	2,680	2,308	86.18
Massachusetts	3,064	2,653	85.67
Michigan	11,441	9,207	79.39
Minnesota	2,483	2,160	85.99
Mississippi	2,553	2,087	81.96

Table 7.9 2012 Screening Results—Eligibility Rate, by State (Weighted Percentages) (continued)

State	SDUs	Eligible DUs	% Eligible DUs
Missouri	2,879	2,409	83.62
Montana	3,295	2,610	78.09
Nebraska	2,556	2,175	85.07
Nevada	2,354	1,879	79.87
New Hampshire	2,990	2,507	83.88
New Jersey	2,622	2,227	84.91
New Mexico	2,771	2,052	73.39
New York	14,547	12,547	85.42
North Carolina	2,848	2,246	76.21
North Dakota	3,374	2,633	77.65
Ohio	11,722	10,122	86.35
Oklahoma	2,960	2,382	79.51
Oregon	2,547	2,250	88.49
Pennsylvania	11,907	10,256	85.02
Rhode Island	2,620	2,190	83.68
South Carolina	3,306	2,666	80.44
South Dakota	2,636	2,163	81.98
Tennessee	2,532	2,095	83.01
Texas	9,048	7,651	84.75
Utah	1,793	1,558	86.99
Vermont	3,292	2,637	78.85
Virginia	2,576	2,293	88.97
Washington	2,700	2,306	85.67
West Virginia	3,222	2,675	82.94
Wisconsin	2,440	2,041	83.27
Wyoming	3,109	2,425	77.59

DUs = dwelling units, SDUs = sample dwelling units.

Table 7.10 2012 Screening Results—Completion Rate, by State (Unweighted Percentages)

State	Eligible DUs	Complete DUs	% Complete DUs
Total	178,586	153,873	86.16
Alabama	2,372	2,141	90.26
Alaska	1,869	1,642	87.85
Arizona	2,143	1,928	89.97
Arkansas	2,292	2,090	91.19
California	8,314	6,852	82.42
Colorado	2,579	2,201	85.34
Connecticut	2,535	2,107	83.12
Delaware	2,292	2,008	87.61
District of Columbia	4,104	3,327	81.07
Florida	10,055	8,516	84.69
Georgia	2,042	1,796	87.95
Hawaii	2,761	2,239	81.09
Idaho	1,939	1,821	93.91
Illinois	9,964	7,678	77.06
Indiana	2,110	1,921	91.04
Iowa	2,199	2,022	91.95
Kansas	2,198	1,977	89.95
Kentucky	2,407	2,202	91.48
Louisiana	2,143	1,977	92.25
Maine	2,858	2,585	90.45
Maryland	2,308	1,802	78.08
Massachusetts	2,653	2,208	83.23
Michigan	9,207	7,826	85.00
Minnesota	2,160	1,975	91.44
Mississippi	2,087	1,951	93.48

Table 7.10 2012 Screening Results—Completion Rate, by State (Unweighted Percentages) (continued)

State	Eligible DUs	Complete DUs	% Complete DUs
Missouri	2,409	2,188	90.83
Montana	2,610	2,415	92.53
Nebraska	2,175	2,018	92.78
Nevada	1,879	1,721	91.59
New Hampshire	2,507	2,191	87.40
New Jersey	2,227	1,935	86.89
New Mexico	2,052	1,889	92.06
New York	12,547	9,115	72.65
North Carolina	2,246	1,990	88.60
North Dakota	2,633	2,461	93.47
Ohio	10,122	9,023	89.14
Oklahoma	2,382	2,173	91.23
Oregon	2,250	2,019	89.73
Pennsylvania	10,256	8,453	82.42
Rhode Island	2,190	1,957	89.36
South Carolina	2,666	2,374	89.05
South Dakota	2,163	2,031	93.90
Tennessee	2,095	1,929	92.08
Texas	7,651	6,792	88.77
Utah	1,558	1,474	94.61
Vermont	2,637	2,317	87.86
Virginia	2,293	2,027	88.40
Washington	2,306	2,078	90.11
West Virginia	2,675	2,399	89.68
Wisconsin	2,041	1,890	92.60
Wyoming	2,425	2,222	91.63

DUs = dwelling units.

Table 7.11 2012 Screening Results—Completion Rate, by State (Weighted Percentages)

State	Eligible DUs	Complete DUs	% Complete DUs
Total	178,586	153,873	86.07
Alabama	2,372	2,141	90.30
Alaska	1,869	1,642	87.82
Arizona	2,143	1,928	90.16
Arkansas	2,292	2,090	90.92
California	8,314	6,852	82.37
Colorado	2,579	2,201	85.23
Connecticut	2,535	2,107	82.76
Delaware	2,292	2,008	87.57
District of Columbia	4,104	3,327	80.90
Florida	10,055	8,516	84.67
Georgia	2,042	1,796	87.94
Hawaii	2,761	2,239	80.80
Idaho	1,939	1,821	93.92
Illinois	9,964	7,678	77.04
Indiana	2,110	1,921	91.01
Iowa	2,199	2,022	91.72
Kansas	2,198	1,977	89.98
Kentucky	2,407	2,202	91.46
Louisiana	2,143	1,977	92.28
Maine	2,858	2,585	90.56
Maryland	2,308	1,802	78.13
Massachusetts	2,653	2,208	83.22
Michigan	9,207	7,826	85.05
Minnesota	2,160	1,975	91.57
Mississippi	2,087	1,951	93.50

Table 7.11 2012 Screening Results—Completion Rate, by State (Weighted Percentages) (continued)

State	Eligible DUs	Complete DUs	% Complete DUs
Missouri	2,409	2,188	90.88
Montana	2,610	2,415	92.62
Nebraska	2,175	2,018	92.74
Nevada	1,879	1,721	91.75
New Hampshire	2,507	2,191	87.40
New Jersey	2,227	1,935	86.87
New Mexico	2,052	1,889	92.22
New York	12,547	9,115	71.89
North Carolina	2,246	1,990	88.48
North Dakota	2,633	2,461	93.42
Ohio	10,122	9,023	89.14
Oklahoma	2,382	2,173	91.22
Oregon	2,250	2,019	89.57
Pennsylvania	10,256	8,453	82.09
Rhode Island	2,190	1,957	89.37
South Carolina	2,666	2,374	88.97
South Dakota	2,163	2,031	93.92
Tennessee	2,095	1,929	91.91
Texas	7,651	6,792	88.52
Utah	1,558	1,474	94.67
Vermont	2,637	2,317	87.81
Virginia	2,293	2,027	88.47
Washington	2,306	2,078	90.10
West Virginia	2,675	2,399	89.39
Wisconsin	2,041	1,890	92.37
Wyoming	2,425	2,222	91.72

DUs = dwelling units.

Table 7.12 2012 Screening Results—Nonresponse Rate, by State (Unweighted Percentages)

State	% Total NR	% Not at Home	% Refused
Total	13.84	1.70	9.41
Alabama	9.74	0.84	7.67
Alaska	12.15	1.44	8.13
Arizona	10.03	1.03	6.67
Arkansas	8.81	0.35	7.77
California	17.58	1.20	12.47
Colorado	14.66	0.85	9.07
Connecticut	16.88	0.79	11.36
Delaware	12.39	1.61	10.25
District of Columbia	18.93	4.63	11.62
Florida	15.31	1.12	10.21
Georgia	12.05	1.08	7.69
Hawaii	18.91	1.99	9.92
Idaho	6.09	0.36	5.36
Illinois	22.94	4.15	13.78
Indiana	8.96	1.28	7.25
Iowa	8.05	1.18	5.64
Kansas	10.05	1.05	7.64
Kentucky	8.52	1.41	5.03
Louisiana	7.75	0.65	7.00
Maine	9.55	1.08	7.42
Maryland	21.92	2.95	15.68
Massachusetts	16.77	1.43	9.46
Michigan	15.00	1.77	12.13
Minnesota	8.56	0.65	7.13
Mississippi	6.52	1.01	4.36

Table 7.12 2012 Screening Results—Nonresponse Rate, by State (Unweighted Percentages) (continued)

State	% Total NR	% Not at Home	% Refused
Missouri	9.17	0.91	7.80
Montana	7.47	0.38	6.32
Nebraska	7.22	0.64	5.01
Nevada	8.41	0.43	7.88
New Hampshire	12.60	0.48	11.65
New Jersey	13.11	2.38	9.07
New Mexico	7.94	0.34	5.70
New York	27.35	3.69	16.12
North Carolina	11.40	1.38	8.15
North Dakota	6.53	0.72	5.66
Ohio	10.86	1.99	7.70
Oklahoma	8.77	0.55	5.75
Oregon	10.27	0.93	6.80
Pennsylvania	17.58	1.99	11.27
Rhode Island	10.64	0.68	7.99
South Carolina	10.95	0.83	8.06
South Dakota	6.10	0.65	4.53
Tennessee	7.92	0.95	6.35
Texas	11.23	2.47	5.99
Utah	5.39	1.41	3.40
Vermont	12.14	1.25	8.68
Virginia	11.60	1.66	8.33
Washington	9.89	0.91	7.94
West Virginia	10.32	1.38	7.25
Wisconsin	7.40	0.20	7.10
Wyoming	8.37	0.82	6.93

NR = nonresponse.

Table 7.13 2012 Screening Results—Nonresponse Rate, by State (Weighted Percentages)

State	% Total NR	% Not at Home	% Refused
Total	13.93	1.62	9.39
Alabama	9.70	0.85	7.70
Alaska	12.18	1.50	8.17
Arizona	9.84	0.91	6.77
Arkansas	9.08	0.34	8.08
California	17.63	1.22	12.47
Colorado	14.77	0.79	9.90
Connecticut	17.24	0.84	11.27
Delaware	12.43	1.62	10.30
District of Columbia	19.10	4.47	11.80
Florida	15.33	1.16	10.23
Georgia	12.06	1.12	7.57
Hawaii	19.20	1.90	9.88
Idaho	6.08	0.36	5.43
Illinois	22.96	4.08	13.86
Indiana	8.99	1.24	7.28
Iowa	8.28	1.18	5.85
Kansas	10.02	1.04	7.70
Kentucky	8.54	1.47	4.92
Louisiana	7.72	0.65	6.99
Maine	9.44	1.10	7.34
Maryland	21.87	2.94	15.68
Massachusetts	16.78	1.45	9.40
Michigan	14.95	1.72	12.10
Minnesota	8.43	0.63	6.92
Mississippi	6.50	1.04	4.33

Table 7.13 2012 Screening Results—Nonresponse Rate, by State (Weighted Percentages) (continued)

State	% Total NR	% Not at Home	% Refused
Missouri	9.12	0.93	7.74
Montana	7.38	0.37	6.27
Nebraska	7.26	0.66	4.96
Nevada	8.25	0.45	7.68
New Hampshire	12.60	0.56	11.57
New Jersey	13.13	2.25	9.11
New Mexico	7.78	0.35	5.73
New York	28.11	3.63	15.95
North Carolina	11.52	1.35	8.14
North Dakota	6.58	0.71	5.71
Ohio	10.86	2.00	7.70
Oklahoma	8.78	0.56	5.77
Oregon	10.43	0.90	6.86
Pennsylvania	17.91	2.02	11.39
Rhode Island	10.63	0.65	8.10
South Carolina	11.03	0.86	8.08
South Dakota	6.08	0.67	4.53
Tennessee	8.09	0.92	6.46
Texas	11.48	2.41	6.13
Utah	5.33	1.40	3.42
Vermont	12.19	1.27	8.74
Virginia	11.53	1.66	8.35
Washington	9.90	0.88	7.94
West Virginia	10.61	1.45	7.26
Wisconsin	7.63	0.18	7.33
Wyoming	8.28	0.83	6.81

NR = nonresponse.

Table 7.14 2012 Screening Refusal Results (Total United States) (Unweighted Percentages)

	Total	
	Count	**%**
Refusal Cases	16,807	100.00
Nothing in it for me	11,152	66.35
No time	1,429	8.50
Government/surveys too invasive	2,395	14.25
Gatekeeper/household member won't allow participation	277	1.65
Confidentiality or survey legitimacy concerns	1,163	6.92
House too messy/too ill	147	0.87
Other	241	1.43
Missing	3	0.02

Table 7.15 2012 Screening Refusal Results (Total United States) (Weighted Percentages)

	Total	
	Count	**%**
Refusal Cases	16,807	100.00
Nothing in it for me	11,152	67.02
No time	1,429	7.64
Government/surveys too invasive	2,395	14.25
Gatekeeper/household member won't allow participation	277	1.77
Confidentiality or survey legitimacy concerns	1,163	7.00
House too messy/too ill	147	0.83
Other	241	1.49
Missing	3	0.01

Table 7.14 2012 Screening Refusal Results (Alabama) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	182	100.00
Nothing in it for me	121	66.48
No time	14	7.69
Government/surveys too invasive	28	15.38
Gatekeeper/household member won't allow participation	2	1.10
Confidentiality or survey legitimacy concerns	8	4.40
House too messy/too ill	2	1.10
Other	7	3.85
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Alabama) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	182	100.00
Nothing in it for me	121	65.96
No time	14	7.30
Government/surveys too invasive	28	16.43
Gatekeeper/household member won't allow participation	2	1.12
Confidentiality or survey legitimacy concerns	8	4.79
House too messy/too ill	2	0.94
Other	7	3.47
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Alaska) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	152	100.00
Nothing in it for me	110	72.37
No time	8	5.26
Government/surveys too invasive	24	15.79
Gatekeeper/household member won't allow participation	1	0.66
Confidentiality or survey legitimacy concerns	7	4.61
House too messy/too ill	1	0.66
Other	1	0.66
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Alaska) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	152	100.00
Nothing in it for me	110	72.27
No time	8	5.60
Government/surveys too invasive	24	15.31
Gatekeeper/household member won't allow participation	1	0.60
Confidentiality or survey legitimacy concerns	7	4.86
House too messy/too ill	1	0.77
Other	1	0.60
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Arizona) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	143	100.00
Nothing in it for me	98	68.53
No time	4	2.80
Government/surveys too invasive	31	21.68
Gatekeeper/household member won't allow participation	1	0.70
Confidentiality or survey legitimacy concerns	6	4.20
House too messy/too ill	1	0.70
Other	2	1.40
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Arizona) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	143	100.00
Nothing in it for me	98	72.45
No time	4	2.34
Government/surveys too invasive	31	19.29
Gatekeeper/household member won't allow participation	1	0.82
Confidentiality or survey legitimacy concerns	6	3.44
House too messy/too ill	1	0.55
Other	2	1.10
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Arkansas) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	178	100.00
Nothing in it for me	134	75.28
No time	11	6.18
Government/surveys too invasive	22	12.36
Gatekeeper/household member won't allow participation	3	1.69
Confidentiality or survey legitimacy concerns	7	3.93
House too messy/too ill	0	0.00
Other	1	0.56
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Arkansas) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	178	100.00
Nothing in it for me	134	75.68
No time	11	6.05
Government/surveys too invasive	22	12.03
Gatekeeper/household member won't allow participation	3	1.63
Confidentiality or survey legitimacy concerns	7	4.02
House too messy/too ill	0	0.00
Other	1	0.59
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (California) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	1,037	100.00
Nothing in it for me	753	72.61
No time	40	3.86
Government/surveys too invasive	143	13.79
Gatekeeper/household member won't allow participation	1	0.10
Confidentiality or survey legitimacy concerns	77	7.43
House too messy/too ill	4	0.39
Other	19	1.83
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (California) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	1,037	100.00
Nothing in it for me	753	72.50
No time	40	3.85
Government/surveys too invasive	143	13.75
Gatekeeper/household member won't allow participation	1	0.11
Confidentiality or survey legitimacy concerns	77	7.52
House too messy/too ill	4	0.39
Other	19	1.88
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Colorado) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	234	100.00
Nothing in it for me	159	67.95
No time	17	7.26
Government/surveys too invasive	44	18.80
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	11	4.70
House too messy/too ill	1	0.43
Other	2	0.85
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Colorado) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	234	100.00
Nothing in it for me	159	62.45
No time	17	7.18
Government/surveys too invasive	44	21.60
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	11	7.70
House too messy/too ill	1	0.35
Other	2	0.73
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Connecticut) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	288	100.00
Nothing in it for me	178	61.81
No time	30	10.42
Government/surveys too invasive	35	12.15
Gatekeeper/household member won't allow participation	7	2.43
Confidentiality or survey legitimacy concerns	33	11.46
House too messy/too ill	2	0.69
Other	3	1.04
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Connecticut) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	288	100.00
Nothing in it for me	178	62.44
No time	30	10.62
Government/surveys too invasive	35	11.62
Gatekeeper/household member won't allow participation	7	2.32
Confidentiality or survey legitimacy concerns	33	11.22
House too messy/too ill	2	0.73
Other	3	1.05
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Delaware) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	235	100.00
Nothing in it for me	129	54.89
No time	19	8.09
Government/surveys too invasive	47	20.00
Gatekeeper/household member won't allow participation	3	1.28
Confidentiality or survey legitimacy concerns	35	14.89
House too messy/too ill	1	0.43
Other	1	0.43
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Delaware) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	235	100.00
Nothing in it for me	129	54.75
No time	19	8.04
Government/surveys too invasive	47	20.26
Gatekeeper/household member won't allow participation	3	1.22
Confidentiality or survey legitimacy concerns	35	14.94
House too messy/too ill	1	0.39
Other	1	0.40
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (District of Columbia) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	477	100.00
Nothing in it for me	321	67.30
No time	79	16.56
Government/surveys too invasive	58	12.16
Gatekeeper/household member won't allow participation	2	0.42
Confidentiality or survey legitimacy concerns	15	3.14
House too messy/too ill	2	0.42
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (District of Columbia) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	477	100.00
Nothing in it for me	321	66.60
No time	79	16.80
Government/surveys too invasive	58	12.71
Gatekeeper/household member won't allow participation	2	0.42
Confidentiality or survey legitimacy concerns	15	3.08
House too messy/too ill	2	0.39
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Florida) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	1,027	100.00
Nothing in it for me	625	60.86
No time	121	11.78
Government/surveys too invasive	138	13.44
Gatekeeper/household member won't allow participation	10	0.97
Confidentiality or survey legitimacy concerns	110	10.71
House too messy/too ill	4	0.39
Other	19	1.85
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Florida) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	1,027	100.00
Nothing in it for me	625	60.36
No time	121	11.43
Government/surveys too invasive	138	14.09
Gatekeeper/household member won't allow participation	10	1.00
Confidentiality or survey legitimacy concerns	110	11.04
House too messy/too ill	4	0.43
Other	19	1.65
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Georgia) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	157	100.00
Nothing in it for me	62	39.49
No time	19	12.10
Government/surveys too invasive	32	20.38
Gatekeeper/household member won't allow participation	4	2.55
Confidentiality or survey legitimacy concerns	34	21.66
House too messy/too ill	0	0.00
Other	6	3.82
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Georgia) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	157	100.00
Nothing in it for me	62	40.54
No time	19	11.26
Government/surveys too invasive	32	19.99
Gatekeeper/household member won't allow participation	4	2.85
Confidentiality or survey legitimacy concerns	34	21.58
House too messy/too ill	0	0.00
Other	6	3.78
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Hawaii) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	274	100.00
Nothing in it for me	197	71.90
No time	40	14.60
Government/surveys too invasive	7	2.55
Gatekeeper/household member won't allow participation	1	0.36
Confidentiality or survey legitimacy concerns	18	6.57
House too messy/too ill	2	0.73
Other	9	3.28
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Hawaii) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	274	100.00
Nothing in it for me	197	71.62
No time	40	14.34
Government/surveys too invasive	7	2.76
Gatekeeper/household member won't allow participation	1	0.31
Confidentiality or survey legitimacy concerns	18	6.94
House too messy/too ill	2	0.69
Other	9	3.35
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Idaho) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	104	100.00
Nothing in it for me	73	70.19
No time	7	6.73
Government/surveys too invasive	21	20.19
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	2	1.92
House too messy/too ill	0	0.00
Other	1	0.96
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Idaho) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	104	100.00
Nothing in it for me	73	70.84
No time	7	6.83
Government/surveys too invasive	21	18.85
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	2	1.82
House too messy/too ill	0	0.00
Other	1	1.66
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Illinois) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	1,373	100.00
Nothing in it for me	1,021	74.36
No time	153	11.14
Government/surveys too invasive	107	7.79
Gatekeeper/household member won't allow participation	15	1.09
Confidentiality or survey legitimacy concerns	59	4.30
House too messy/too ill	6	0.44
Other	12	0.87
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Illinois) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	1,373	100.00
Nothing in it for me	1,021	74.29
No time	153	11.10
Government/surveys too invasive	107	7.94
Gatekeeper/household member won't allow participation	15	1.13
Confidentiality or survey legitimacy concerns	59	4.25
House too messy/too ill	6	0.42
Other	12	0.87
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Indiana) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	153	100.00
Nothing in it for me	71	46.41
No time	12	7.84
Government/surveys too invasive	55	35.95
Gatekeeper/household member won't allow participation	3	1.96
Confidentiality or survey legitimacy concerns	10	6.54
House too messy/too ill	1	0.65
Other	1	0.65
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Indiana) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	153	100.00
Nothing in it for me	71	45.80
No time	12	7.95
Government/surveys too invasive	55	36.57
Gatekeeper/household member won't allow participation	3	1.82
Confidentiality or survey legitimacy concerns	10	6.59
House too messy/too ill	1	0.72
Other	1	0.55
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Iowa) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	124	100.00
Nothing in it for me	78	62.90
No time	3	2.42
Government/surveys too invasive	28	22.58
Gatekeeper/household member won't allow participation	8	6.45
Confidentiality or survey legitimacy concerns	5	4.03
House too messy/too ill	1	0.81
Other	1	0.81
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Iowa) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	124	100.00
Nothing in it for me	78	61.45
No time	3	2.41
Government/surveys too invasive	28	23.54
Gatekeeper/household member won't allow participation	8	7.51
Confidentiality or survey legitimacy concerns	5	3.64
House too messy/too ill	1	0.59
Other	1	0.87
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Kansas) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	168	100.00
Nothing in it for me	130	77.38
No time	10	5.95
Government/surveys too invasive	18	10.71
Gatekeeper/household member won't allow participation	3	1.79
Confidentiality or survey legitimacy concerns	7	4.17
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Kansas) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	168	100.00
Nothing in it for me	130	77.57
No time	10	5.81
Government/surveys too invasive	18	10.49
Gatekeeper/household member won't allow participation	3	2.03
Confidentiality or survey legitimacy concerns	7	4.10
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Kentucky) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	121	100.00
Nothing in it for me	73	60.33
No time	15	12.40
Government/surveys too invasive	18	14.88
Gatekeeper/household member won't allow participation	2	1.65
Confidentiality or survey legitimacy concerns	13	10.74
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Kentucky) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	121	100.00
Nothing in it for me	73	60.94
No time	15	11.56
Government/surveys too invasive	18	15.48
Gatekeeper/household member won't allow participation	2	1.60
Confidentiality or survey legitimacy concerns	13	10.42
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Louisiana) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	150	100.00
Nothing in it for me	112	74.67
No time	13	8.67
Government/surveys too invasive	19	12.67
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	3	2.00
House too messy/too ill	3	2.00
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Louisiana) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	150	100.00
Nothing in it for me	112	75.79
No time	13	9.57
Government/surveys too invasive	19	11.10
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	3	1.65
House too messy/too ill	3	1.89
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Maine) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	212	100.00
Nothing in it for me	142	66.98
No time	23	10.85
Government/surveys too invasive	30	14.15
Gatekeeper/household member won't allow participation	4	1.89
Confidentiality or survey legitimacy concerns	12	5.66
House too messy/too ill	0	0.00
Other	1	0.47
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Maine) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	212	100.00
Nothing in it for me	142	66.52
No time	23	10.99
Government/surveys too invasive	30	14.28
Gatekeeper/household member won't allow participation	4	1.92
Confidentiality or survey legitimacy concerns	12	5.84
House too messy/too ill	0	0.00
Other	1	0.45
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Maryland) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	362	100.00
Nothing in it for me	281	77.62
No time	33	9.12
Government/surveys too invasive	36	9.94
Gatekeeper/household member won't allow participation	4	1.10
Confidentiality or survey legitimacy concerns	2	0.55
House too messy/too ill	4	1.10
Other	2	0.55
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Maryland) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	362	100.00
Nothing in it for me	281	77.20
No time	33	9.27
Government/surveys too invasive	36	10.15
Gatekeeper/household member won't allow participation	4	1.12
Confidentiality or survey legitimacy concerns	2	0.55
House too messy/too ill	4	1.12
Other	2	0.59
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Massachusetts) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	251	100.00
Nothing in it for me	145	57.77
No time	22	8.76
Government/surveys too invasive	49	19.52
Gatekeeper/household member won't allow participation	4	1.59
Confidentiality or survey legitimacy concerns	22	8.76
House too messy/too ill	2	0.80
Other	7	2.79
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Massachusetts) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	251	100.00
Nothing in it for me	145	57.59
No time	22	8.73
Government/surveys too invasive	49	19.71
Gatekeeper/household member won't allow participation	4	1.60
Confidentiality or survey legitimacy concerns	22	8.78
House too messy/too ill	2	0.73
Other	7	2.86
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Michigan) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	1,117	100.00
Nothing in it for me	836	74.84
No time	76	6.80
Government/surveys too invasive	110	9.85
Gatekeeper/household member won't allow participation	12	1.07
Confidentiality or survey legitimacy concerns	31	2.78
House too messy/too ill	38	3.40
Other	14	1.25
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Michigan) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	1,117	100.00
Nothing in it for me	836	74.79
No time	76	6.85
Government/surveys too invasive	110	9.99
Gatekeeper/household member won't allow participation	12	1.05
Confidentiality or survey legitimacy concerns	31	2.86
House too messy/too ill	38	3.27
Other	14	1.19
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Minnesota) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	154	100.00
Nothing in it for me	89	57.79
No time	4	2.60
Government/surveys too invasive	41	26.62
Gatekeeper/household member won't allow participation	3	1.95
Confidentiality or survey legitimacy concerns	13	8.44
House too messy/too ill	0	0.00
Other	4	2.60
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Minnesota) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	154	100.00
Nothing in it for me	89	58.48
No time	4	2.19
Government/surveys too invasive	41	26.56
Gatekeeper/household member won't allow participation	3	1.99
Confidentiality or survey legitimacy concerns	13	8.41
House too messy/too ill	0	0.00
Other	4	2.37
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Mississippi) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	91	100.00
Nothing in it for me	71	78.02
No time	4	4.40
Government/surveys too invasive	8	8.79
Gatekeeper/household member won't allow participation	7	7.69
Confidentiality or survey legitimacy concerns	0	0.00
House too messy/too ill	0	0.00
Other	1	1.10
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Mississippi) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	91	100.00
Nothing in it for me	71	78.22
No time	4	4.27
Government/surveys too invasive	8	8.68
Gatekeeper/household member won't allow participation	7	7.59
Confidentiality or survey legitimacy concerns	0	0.00
House too messy/too ill	0	0.00
Other	1	1.24
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Missouri) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	188	100.00
Nothing in it for me	125	66.49
No time	10	5.32
Government/surveys too invasive	28	14.89
Gatekeeper/household member won't allow participation	5	2.66
Confidentiality or survey legitimacy concerns	17	9.04
House too messy/too ill	2	1.06
Other	1	0.53
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Missouri) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	188	100.00
Nothing in it for me	125	66.17
No time	10	5.24
Government/surveys too invasive	28	15.01
Gatekeeper/household member won't allow participation	5	2.57
Confidentiality or survey legitimacy concerns	17	9.31
House too messy/too ill	2	1.21
Other	1	0.49
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Montana) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	165	100.00
Nothing in it for me	113	68.48
No time	8	4.85
Government/surveys too invasive	30	18.18
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	10	6.06
House too messy/too ill	0	0.00
Other	4	2.42
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Montana) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	165	100.00
Nothing in it for me	113	68.11
No time	8	4.73
Government/surveys too invasive	30	18.56
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	10	6.20
House too messy/too ill	0	0.00
Other	4	2.40
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Nebraska) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	109	100.00
Nothing in it for me	55	50.46
No time	11	10.09
Government/surveys too invasive	33	30.28
Gatekeeper/household member won't allow participation	1	0.92
Confidentiality or survey legitimacy concerns	7	6.42
House too messy/too ill	1	0.92
Other	1	0.92
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Nebraska) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	109	100.00
Nothing in it for me	55	50.19
No time	11	9.87
Government/surveys too invasive	33	30.61
Gatekeeper/household member won't allow participation	1	1.00
Confidentiality or survey legitimacy concerns	7	6.60
House too messy/too ill	1	0.89
Other	1	0.86
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Nevada) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	148	100.00
Nothing in it for me	89	60.14
No time	10	6.76
Government/surveys too invasive	25	16.89
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	9	6.08
House too messy/too ill	0	0.00
Other	15	10.14
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Nevada) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	148	100.00
Nothing in it for me	89	59.45
No time	10	6.68
Government/surveys too invasive	25	16.80
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	9	6.44
House too messy/too ill	0	0.00
Other	15	10.64
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (New Hampshire) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	292	100.00
Nothing in it for me	222	76.03
No time	11	3.77
Government/surveys too invasive	35	11.99
Gatekeeper/household member won't allow participation	3	1.03
Confidentiality or survey legitimacy concerns	16	5.48
House too messy/too ill	1	0.34
Other	4	1.37
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (New Hampshire) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	292	100.00
Nothing in it for me	222	75.94
No time	11	3.76
Government/surveys too invasive	35	11.96
Gatekeeper/household member won't allow participation	3	1.03
Confidentiality or survey legitimacy concerns	16	5.55
House too messy/too ill	1	0.37
Other	4	1.39
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (New Jersey) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	202	100.00
Nothing in it for me	160	79.21
No time	24	11.88
Government/surveys too invasive	7	3.47
Gatekeeper/household member won't allow participation	2	0.99
Confidentiality or survey legitimacy concerns	4	1.98
House too messy/too ill	3	1.49
Other	2	0.99
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (New Jersey) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	202	100.00
Nothing in it for me	160	79.23
No time	24	11.86
Government/surveys too invasive	7	3.27
Gatekeeper/household member won't allow participation	2	1.18
Confidentiality or survey legitimacy concerns	4	1.99
House too messy/too ill	3	1.44
Other	2	1.03
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (New Mexico) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	117	100.00
Nothing in it for me	75	64.10
No time	10	8.55
Government/surveys too invasive	23	19.66
Gatekeeper/household member won't allow participation	3	2.56
Confidentiality or survey legitimacy concerns	6	5.13
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (New Mexico) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	117	100.00
Nothing in it for me	75	65.25
No time	10	8.35
Government/surveys too invasive	23	18.84
Gatekeeper/household member won't allow participation	3	2.81
Confidentiality or survey legitimacy concerns	6	4.74
House too messy/too ill	0	0.00
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (New York) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	2,023	100.00
Nothing in it for me	1,416	70.00
No time	168	8.30
Government/surveys too invasive	202	9.99
Gatekeeper/household member won't allow participation	73	3.61
Confidentiality or survey legitimacy concerns	136	6.72
House too messy/too ill	23	1.14
Other	5	0.25
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (New York) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	2,023	100.00
Nothing in it for me	1,416	70.00
No time	168	8.25
Government/surveys too invasive	202	10.01
Gatekeeper/household member won't allow participation	73	3.69
Confidentiality or survey legitimacy concerns	136	6.69
House too messy/too ill	23	1.12
Other	5	0.24
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (North Carolina) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	183	100.00
Nothing in it for me	102	55.74
No time	27	14.75
Government/surveys too invasive	28	15.30
Gatekeeper/household member won't allow participation	6	3.28
Confidentiality or survey legitimacy concerns	13	7.10
House too messy/too ill	2	1.09
Other	5	2.73
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (North Carolina) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	183	100.00
Nothing in it for me	102	55.72
No time	27	14.42
Government/surveys too invasive	28	15.39
Gatekeeper/household member won't allow participation	6	3.17
Confidentiality or survey legitimacy concerns	13	7.31
House too messy/too ill	2	1.09
Other	5	2.88
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (North Dakota) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	149	100.00
Nothing in it for me	31	20.81
No time	33	22.15
Government/surveys too invasive	58	38.93
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	22	14.77
House too messy/too ill	1	0.67
Other	4	2.68
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (North Dakota) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	149	100.00
Nothing in it for me	31	20.92
No time	33	22.95
Government/surveys too invasive	58	37.97
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	22	14.36
House too messy/too ill	1	0.76
Other	4	3.04
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Ohio) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	779	100.00
Nothing in it for me	561	72.02
No time	52	6.68
Government/surveys too invasive	122	15.66
Gatekeeper/household member won't allow participation	5	0.64
Confidentiality or survey legitimacy concerns	33	4.24
House too messy/too ill	4	0.51
Other	2	0.26
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Ohio) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	779	100.00
Nothing in it for me	561	71.88
No time	52	6.79
Government/surveys too invasive	122	15.77
Gatekeeper/household member won't allow participation	5	0.68
Confidentiality or survey legitimacy concerns	33	4.12
House too messy/too ill	4	0.50
Other	2	0.27
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Oklahoma) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	137	100.00
Nothing in it for me	113	82.48
No time	8	5.84
Government/surveys too invasive	9	6.57
Gatekeeper/household member won't allow participation	2	1.46
Confidentiality or survey legitimacy concerns	4	2.92
House too messy/too ill	1	0.73
Other	0	0.00
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Oklahoma) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	137	100.00
Nothing in it for me	113	82.14
No time	8	5.94
Government/surveys too invasive	9	6.74
Gatekeeper/household member won't allow participation	2	1.43
Confidentiality or survey legitimacy concerns	4	2.86
House too messy/too ill	1	0.89
Other	0	0.00
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Oregon) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	153	100.00
Nothing in it for me	112	73.20
No time	2	1.31
Government/surveys too invasive	14	9.15
Gatekeeper/household member won't allow participation	2	1.31
Confidentiality or survey legitimacy concerns	18	11.76
House too messy/too ill	2	1.31
Other	3	1.96
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Oregon) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	153	100.00
Nothing in it for me	112	73.43
No time	2	1.24
Government/surveys too invasive	14	8.73
Gatekeeper/household member won't allow participation	2	1.34
Confidentiality or survey legitimacy concerns	18	12.28
House too messy/too ill	2	1.16
Other	3	1.82
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Pennsylvania) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	1,156	100.00
Nothing in it for me	622	53.81
No time	113	9.78
Government/surveys too invasive	172	14.88
Gatekeeper/household member won't allow participation	27	2.34
Confidentiality or survey legitimacy concerns	193	16.70
House too messy/too ill	13	1.12
Other	14	1.21
Missing	2	0.17

Table 7.15 2012 Screening Refusal Results (Pennsylvania) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	1,156	100.00
Nothing in it for me	622	53.98
No time	113	10.30
Government/surveys too invasive	172	14.99
Gatekeeper/household member won't allow participation	27	2.43
Confidentiality or survey legitimacy concerns	193	15.81
House too messy/too ill	13	1.17
Other	14	1.15
Missing	2	0.18

Table 7.14 2012 Screening Refusal Results (Rhode Island) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	175	100.00
Nothing in it for me	95	54.29
No time	11	6.29
Government/surveys too invasive	32	18.29
Gatekeeper/household member won't allow participation	6	3.43
Confidentiality or survey legitimacy concerns	28	16.00
House too messy/too ill	2	1.14
Other	1	0.57
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Rhode Island) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	175	100.00
Nothing in it for me	95	55.69
No time	11	6.26
Government/surveys too invasive	32	17.00
Gatekeeper/household member won't allow participation	6	3.37
Confidentiality or survey legitimacy concerns	28	15.93
House too messy/too ill	2	1.10
Other	1	0.64
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (South Carolina) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	215	100.00
Nothing in it for me	127	59.07
No time	14	6.51
Government/surveys too invasive	53	24.65
Gatekeeper/household member won't allow participation	7	3.26
Confidentiality or survey legitimacy concerns	3	1.40
House too messy/too ill	1	0.47
Other	10	4.65
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (South Carolina) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	215	100.00
Nothing in it for me	127	58.11
No time	14	7.13
Government/surveys too invasive	53	24.95
Gatekeeper/household member won't allow participation	7	3.26
Confidentiality or survey legitimacy concerns	3	1.35
House too messy/too ill	1	0.49
Other	10	4.70
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (South Dakota) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	98	100.00
Nothing in it for me	55	56.12
No time	4	4.08
Government/surveys too invasive	31	31.63
Gatekeeper/household member won't allow participation	1	1.02
Confidentiality or survey legitimacy concerns	5	5.10
House too messy/too ill	1	1.02
Other	1	1.02
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (South Dakota) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	98	100.00
Nothing in it for me	55	56.76
No time	4	4.43
Government/surveys too invasive	31	31.08
Gatekeeper/household member won't allow participation	1	0.96
Confidentiality or survey legitimacy concerns	5	5.14
House too messy/too ill	1	0.81
Other	1	0.82
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Tennessee) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	133	100.00
Nothing in it for me	107	80.45
No time	9	6.77
Government/surveys too invasive	12	9.02
Gatekeeper/household member won't allow participation	1	0.75
Confidentiality or survey legitimacy concerns	1	0.75
House too messy/too ill	0	0.00
Other	3	2.26
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Tennessee) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	133	100.00
Nothing in it for me	107	80.02
No time	9	6.54
Government/surveys too invasive	12	9.39
Gatekeeper/household member won't allow participation	1	0.81
Confidentiality or survey legitimacy concerns	1	0.84
House too messy/too ill	0	0.00
Other	3	2.40
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Texas) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	458	100.00
Nothing in it for me	308	67.25
No time	20	4.37
Government/surveys too invasive	69	15.07
Gatekeeper/household member won't allow participation	12	2.62
Confidentiality or survey legitimacy concerns	27	5.90
House too messy/too ill	3	0.66
Other	19	4.15
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Texas) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	458	100.00
Nothing in it for me	308	66.73
No time	20	4.16
Government/surveys too invasive	69	16.14
Gatekeeper/household member won't allow participation	12	2.83
Confidentiality or survey legitimacy concerns	27	5.67
House too messy/too ill	3	0.63
Other	19	3.84
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Utah) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	53	100.00
Nothing in it for me	34	64.15
No time	3	5.66
Government/surveys too invasive	13	24.53
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	2	3.77
House too messy/too ill	0	0.00
Other	0	0.00
Missing	1	1.89

Table 7.15 2012 Screening Refusal Results (Utah) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	53	100.00
Nothing in it for me	34	65.35
No time	3	5.23
Government/surveys too invasive	13	23.61
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	2	4.03
House too messy/too ill	0	0.00
Other	0	0.00
Missing	1	1.77

Table 7.14 2012 Screening Refusal Results (Vermont) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	229	100.00
Nothing in it for me	79	34.50
No time	44	19.21
Government/surveys too invasive	87	37.99
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	10	4.37
House too messy/too ill	3	1.31
Other	6	2.62
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Vermont) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	229	100.00
Nothing in it for me	79	35.08
No time	44	19.48
Government/surveys too invasive	87	37.21
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	10	4.60
House too messy/too ill	3	1.14
Other	6	2.49
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Virginia) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	191	100.00
Nothing in it for me	135	70.68
No time	16	8.38
Government/surveys too invasive	24	12.57
Gatekeeper/household member won't allow participation	12	6.28
Confidentiality or survey legitimacy concerns	2	1.05
House too messy/too ill	1	0.52
Other	1	0.52
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Virginia) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	191	100.00
Nothing in it for me	135	70.69
No time	16	8.37
Government/surveys too invasive	24	12.65
Gatekeeper/household member won't allow participation	12	6.15
Confidentiality or survey legitimacy concerns	2	1.05
House too messy/too ill	1	0.53
Other	1	0.55
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Washington) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	183	100.00
Nothing in it for me	105	57.38
No time	7	3.83
Government/surveys too invasive	45	24.59
Gatekeeper/household member won't allow participation	3	1.64
Confidentiality or survey legitimacy concerns	19	10.38
House too messy/too ill	3	1.64
Other	1	0.55
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Washington) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	183	100.00
Nothing in it for me	105	58.17
No time	7	3.96
Government/surveys too invasive	45	23.30
Gatekeeper/household member won't allow participation	3	1.49
Confidentiality or survey legitimacy concerns	19	10.79
House too messy/too ill	3	1.64
Other	1	0.65
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (West Virginia) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	194	100.00
Nothing in it for me	131	67.53
No time	21	10.82
Government/surveys too invasive	25	12.89
Gatekeeper/household member won't allow participation	4	2.06
Confidentiality or survey legitimacy concerns	8	4.12
House too messy/too ill	1	0.52
Other	4	2.06
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (West Virginia) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	194	100.00
Nothing in it for me	131	69.53
No time	21	10.44
Government/surveys too invasive	25	12.02
Gatekeeper/household member won't allow participation	4	1.96
Confidentiality or survey legitimacy concerns	8	3.66
House too messy/too ill	1	0.52
Other	4	1.87
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Wisconsin) (Unweighted Percentages)

	Total	
	Count	%
Refusal Cases	145	100.00
Nothing in it for me	107	73.79
No time	2	1.38
Government/surveys too invasive	22	15.17
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	9	6.21
House too messy/too ill	4	2.76
Other	1	0.69
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Wisconsin) (Weighted Percentages)

	Total	
	Count	%
Refusal Cases	145	100.00
Nothing in it for me	107	74.47
No time	2	1.17
Government/surveys too invasive	22	14.76
Gatekeeper/household member won't allow participation	0	0.00
Confidentiality or survey legitimacy concerns	9	6.17
House too messy/too ill	4	2.66
Other	1	0.78
Missing	0	0.00

Table 7.14 2012 Screening Refusal Results (Wyoming) (Unweighted Percentages)

	Total	
	Count	**%**
Refusal Cases	168	100.00
Nothing in it for me	64	38.10
No time	14	8.33
Government/surveys too invasive	47	27.98
Gatekeeper/household member won't allow participation	2	1.19
Confidentiality or survey legitimacy concerns	21	12.50
House too messy/too ill	0	0.00
Other	20	11.90
Missing	0	0.00

Table 7.15 2012 Screening Refusal Results (Wyoming) (Weighted Percentages)

	Total	
	Count	**%**
Refusal Cases	168	100.00
Nothing in it for me	64	36.93
No time	14	8.32
Government/surveys too invasive	47	28.86
Gatekeeper/household member won't allow participation	2	1.11
Confidentiality or survey legitimacy concerns	21	12.71
House too messy/too ill	0	0.00
Other	20	12.06
Missing	0	0.00

Table 7.16 Interview Response Rates, by Age (Total United States)

	Unweighted		Weighted	
	2011	2012	2011	2012
Age Category				
12-17	84.37	82.85	84.95	82.84
18-25	80.74	79.48	80.48	79.26
26-34	76.01	75.72	75.25	75.33
35-49	74.97	73.31	74.75	73.00
50-64	71.48	70.34	71.34	69.18
65+	66.52	66.41	65.60	65.60

Table 7.17 Interview Response Rates, by Small Age Groups (Total United States)

	Unweighted		Weighted	
	2011	2012	2011	2012
Age Group				
12	82.48	81.95	82.04	81.88
13	85.11	82.40	86.30	82.90
14	85.41	83.26	86.61	82.57
15	84.78	84.24	85.68	84.09
16	84.83	82.87	84.84	82.54
17	83.55	82.44	84.08	83.05
18	83.57	82.97	84.14	83.38
19	83.25	82.04	83.24	82.11
20	82.33	81.50	82.26	80.67
21	81.90	79.52	81.48	80.24
22	80.14	78.66	79.60	78.29
23	78.46	78.64	77.46	78.34
24	79.15	75.96	78.71	74.88
25	77.27	77.11	76.56	76.44
26-29	76.76	75.76	76.21	75.79
30-34	75.38	75.69	74.42	74.96
35-39	76.88	73.85	77.55	73.34
40-44	74.75	73.30	73.83	73.27
45-49	73.49	72.83	73.17	72.44
50-54	72.83	71.72	72.51	70.62
55-59	69.65	69.88	69.71	68.35
60-64	71.83	69.09	71.60	68.19
65-69	69.85	70.93	68.19	69.64
70-74	67.89	69.28	68.43	67.11
75+	63.21	61.55	61.97	61.87

Table 7.18 2012 Interview Results, by Gender and Age (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Male								
Eligible Cases	13,806	100.00	13,947	100.00	15,189	100.00	42,942	100.00
70 - Interview Complete	11,394	82.53	10,891	78.09	10,584	69.68	32,869	76.54
71 - No One at DU*	239	1.73	536	3.84	591	3.89	1,366	3.18
77 - Refusal	463	3.35	2,084	14.94	3,457	22.76	6,004	13.98
Other	1,710	12.39	436	3.13	557	3.67	2,703	6.29
Female								
Eligible Cases	13,341	100.00	14,692	100.00	16,681	100.00	44,714	100.00
70 - Interview Complete	11,098	83.19	11,871	80.80	12,471	74.76	35,440	79.26
71 - No One at DU*	229	1.72	554	3.77	443	2.66	1,226	2.74
77 - Refusal	437	3.28	1,892	12.88	3,155	18.91	5,484	12.26
Other	1,577	11.82	375	2.55	612	3.67	2,564	5.73
Total								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.85	22,762	79.48	23,055	72.34	68,309	77.93
71 - No One at DU*	468	1.72	1,090	3.81	1,034	3.24	2,592	2.96
77 - Refusal	900	3.32	3,976	13.88	6,612	20.75	11,488	13.11
Other	3,287	12.11	811	2.83	1,169	3.67	5,267	6.01

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.19 2012 Interview Results, by Gender and Age (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Male								
Eligible Cases	13,806	100.00	13,947	100.00	15,189	100.00	42,942	100.00
70 - Interview Complete	11,394	82.30	10,891	78.03	10,584	68.54	32,869	71.24
71 - No One at DU*	239	1.71	536	3.52	591	3.86	1,366	3.60
77 - Refusal	463	3.42	2,084	15.13	3,457	22.90	6,004	19.86
Other	1,710	12.57	436	3.32	557	4.69	2,703	5.30
Female								
Eligible Cases	13,341	100.00	14,692	100.00	16,681	100.00	44,714	100.00
70 - Interview Complete	11,098	83.39	11,871	80.49	12,471	72.75	35,440	74.71
71 - No One at DU*	229	1.64	554	3.68	443	2.38	1,226	2.47
77 - Refusal	437	3.32	1,892	13.10	3,155	19.85	5,484	17.48
Other	1,577	11.65	375	2.74	612	5.03	2,564	5.34
Total								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.84	22,762	79.26	23,055	70.76	68,309	73.04
71 - No One at DU*	468	1.68	1,090	3.60	1,034	3.08	2,592	3.01
77 - Refusal	900	3.37	3,976	14.11	6,612	21.30	11,488	18.63
Other	3,287	12.12	811	3.03	1,169	4.87	5,267	5.32

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.20 2012 Interview Results, by Age (Total United States) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.85	22,762	79.48	23,055	72.34	68,309	77.93
71 - No One at DU	192	0.71	492	1.72	463	1.45	1,147	1.31
72 - Respondent Unavailable	276	1.02	598	2.09	571	1.79	1,445	1.65
73 - Break Off (Partial Interview)	0	0.00	8	0.03	13	0.04	21	0.02
74 - Physically/Mentally Incompetent	274	1.01	216	0.75	533	1.67	1,023	1.17
75 - Language Barrier - Spanish	9	0.03	40	0.14	67	0.21	116	0.13
76 - Language Barrier - Other	30	0.11	71	0.25	318	1.00	419	0.48
77 - Refusal	900	3.32	3,976	13.88	6,612	20.75	11,488	13.11
78 - Parental Refusal	2,787	10.27	0	0.00	0	0.00	2,787	3.18
Other	187	0.69	476	1.66	238	0.75	901	1.03

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Total United States) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.84	22,762	79.26	23,055	70.76	68,309	73.04
71 - No One at DU	192	0.67	492	1.53	463	1.29	1,147	1.26
72 - Respondent Unavailable	276	1.00	598	2.07	571	1.79	1,445	1.75
73 - Break Off (Partial Interview)	0	0.00	8	0.03	13	0.06	21	0.05
74 - Physically/Mentally Incompetent	274	1.16	216	0.77	533	2.25	1,023	1.95
75 - Language Barrier - Spanish	9	0.02	40	0.13	67	0.19	116	0.17
76 - Language Barrier - Other	30	0.15	71	0.31	318	1.54	419	1.24
77 - Refusal	900	3.37	3,976	14.11	6,612	21.30	11,488	18.63
78 - Parental Refusal	2,787	10.06	0	0.00	0	0.00	2,787	0.97
Other	187	0.73	476	1.79	238	0.83	901	0.95

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Alabama) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	342	100.00	383	100.00	420	100.00	1,145	100.00
70 - Interview Complete	278	81.29	312	81.46	311	74.05	901	78.69
71 - No One at DU	2	0.58	5	1.31	3	0.71	10	0.87
72 - Respondent Unavailable	8	2.34	12	3.13	13	3.10	33	2.88
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.17	2	0.52	14	3.33	20	1.75
75 - Language Barrier - Spanish	2	0.58	2	0.52	4	0.95	8	0.70
76 - Language Barrier - Other	0	0.00	0	0.00	2	0.48	2	0.17
77 - Refusal	16	4.68	40	10.44	73	17.38	129	11.27
78 - Parental Refusal	31	9.06	0	0.00	0	0.00	31	2.71
Other	1	0.29	10	2.61	0	0.00	11	0.96

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Alabama) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	342	100.00	383	100.00	420	100.00	1,145	100.00
70 - Interview Complete	278	80.41	312	80.90	311	72.65	901	74.57
71 - No One at DU	2	0.51	5	1.15	3	0.43	10	0.54
72 - Respondent Unavailable	8	1.92	12	2.81	13	2.82	33	2.73
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.48	2	0.56	14	3.51	20	2.90
75 - Language Barrier - Spanish	2	0.45	2	0.64	4	0.57	8	0.57
76 - Language Barrier - Other	0	0.00	0	0.00	2	0.58	2	0.44
77 - Refusal	16	4.95	40	11.50	73	19.44	129	16.88
78 - Parental Refusal	31	10.08	0	0.00	0	0.00	31	1.03
Other	1	0.20	10	2.44	0	0.00	11	0.35

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Alaska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	304	100.00	348	100.00	424	100.00	1,076	100.00
70 - Interview Complete	233	76.64	286	82.18	310	73.11	829	77.04
71 - No One at DU	5	1.64	10	2.87	8	1.89	23	2.14
72 - Respondent Unavailable	12	3.95	9	2.59	6	1.42	27	2.51
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.24	1	0.09
74 - Physically/Mentally Incompetent	4	1.32	3	0.86	6	1.42	13	1.21
75 - Language Barrier - Spanish	2	0.66	0	0.00	2	0.47	4	0.37
76 - Language Barrier - Other	0	0.00	1	0.29	4	0.94	5	0.46
77 - Refusal	17	5.59	35	10.06	83	19.58	135	12.55
78 - Parental Refusal	30	9.87	0	0.00	0	0.00	30	2.79
Other	1	0.33	4	1.15	4	0.94	9	0.84

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Alaska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	304	100.00	348	100.00	424	100.00	1,076	100.00
70 - Interview Complete	233	76.07	286	82.25	310	71.44	829	73.34
71 - No One at DU	5	1.84	10	2.47	8	1.91	23	1.98
72 - Respondent Unavailable	12	4.64	9	2.63	6	1.62	27	2.04
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.49	1	0.38
74 - Physically/Mentally Incompetent	4	1.36	3	1.06	6	2.73	13	2.38
75 - Language Barrier - Spanish	2	0.44	0	0.00	2	0.49	4	0.42
76 - Language Barrier - Other	0	0.00	1	0.15	4	0.95	5	0.75
77 - Refusal	17	4.81	35	10.43	83	19.48	135	16.86
78 - Parental Refusal	30	10.48	0	0.00	0	0.00	30	1.00
Other	1	0.37	4	1.01	4	0.88	9	0.85

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Arizona) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	366	100.00	371	100.00	402	100.00	1,139	100.00
70 - Interview Complete	312	85.25	293	78.98	317	78.86	922	80.95
71 - No One at DU	2	0.55	9	2.43	5	1.24	16	1.40
72 - Respondent Unavailable	5	1.37	9	2.43	10	2.49	24	2.11
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.55	1	0.27	7	1.74	10	0.88
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	5	1.37	6	1.62	6	1.49	17	1.49
77 - Refusal	20	5.46	47	12.67	55	13.68	122	10.71
78 - Parental Refusal	15	4.10	0	0.00	0	0.00	15	1.32
Other	5	1.37	6	1.62	2	0.50	13	1.14

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Arizona) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	366	100.00	371	100.00	402	100.00	1,139	100.00
70 - Interview Complete	312	85.61	293	74.97	317	76.39	922	77.11
71 - No One at DU	2	0.49	9	2.02	5	1.70	16	1.62
72 - Respondent Unavailable	5	1.33	9	2.84	10	2.13	24	2.15
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.58	1	0.64	7	1.98	10	1.67
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	5	1.75	6	2.28	6	2.10	17	2.09
77 - Refusal	20	5.49	47	15.73	55	14.87	122	14.06
78 - Parental Refusal	15	3.70	0	0.00	0	0.00	15	0.36
Other	5	1.06	6	1.52	2	0.82	13	0.94

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Arkansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	394	100.00	404	100.00	414	100.00	1,212	100.00
70 - Interview Complete	312	79.19	310	76.73	291	70.29	913	75.33
71 - No One at DU	2	0.51	1	0.25	3	0.72	6	0.50
72 - Respondent Unavailable	0	0.00	4	0.99	4	0.97	8	0.66
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.52	1	0.25	6	1.45	13	1.07
75 - Language Barrier - Spanish	1	0.25	1	0.25	1	0.24	3	0.25
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.24	1	0.08
77 - Refusal	22	5.58	79	19.55	105	25.36	206	17.00
78 - Parental Refusal	48	12.18	0	0.00	0	0.00	48	3.96
Other	3	0.76	8	1.98	3	0.72	14	1.16

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Arkansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	394	100.00	404	100.00	414	100.00	1,212	100.00
70 - Interview Complete	312	78.13	310	75.45	291	67.71	913	69.77
71 - No One at DU	2	0.44	1	0.22	3	0.80	6	0.69
72 - Respondent Unavailable	0	0.00	4	0.92	4	0.80	8	0.74
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.61	1	0.19	6	2.06	13	1.76
75 - Language Barrier - Spanish	1	0.18	1	0.22	1	0.13	3	0.15
76 - Language Barrier - Other	0	0.00	0	0.00	1	1.07	1	0.82
77 - Refusal	22	6.45	79	20.91	105	26.75	206	24.01
78 - Parental Refusal	48	12.57	0	0.00	0	0.00	48	1.20
Other	3	0.63	8	2.09	3	0.68	14	0.87

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (California) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,409	100.00	1,584	100.00	1,786	100.00	4,779	100.00
70 - Interview Complete	1,159	82.26	1,216	76.77	1,233	69.04	3,608	75.50
71 - No One at DU	3	0.21	6	0.38	11	0.62	20	0.42
72 - Respondent Unavailable	8	0.57	14	0.88	17	0.95	39	0.82
73 - Break Off (Partial Interview)	0	0.00	1	0.06	1	0.06	2	0.04
74 - Physically/Mentally Incompetent	11	0.78	14	0.88	31	1.74	56	1.17
75 - Language Barrier - Spanish	1	0.07	0	0.00	0	0.00	1	0.02
76 - Language Barrier - Other	3	0.21	11	0.69	44	2.46	58	1.21
77 - Refusal	70	4.97	296	18.69	425	23.80	791	16.55
78 - Parental Refusal	145	10.29	0	0.00	0	0.00	145	3.03
Other	9	0.64	26	1.64	24	1.34	59	1.23

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (California) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,409	100.00	1,584	100.00	1,786	100.00	4,779	100.00
70 - Interview Complete	1,159	81.82	1,216	76.51	1,233	67.51	3,608	70.20
71 - No One at DU	3	0.22	6	0.36	11	0.57	20	0.51
72 - Respondent Unavailable	8	0.60	14	1.03	17	0.79	39	0.80
73 - Break Off (Partial Interview)	0	0.00	1	0.05	1	0.08	2	0.06
74 - Physically/Mentally Incompetent	11	0.93	14	0.79	31	2.18	56	1.86
75 - Language Barrier - Spanish	1	0.07	0	0.00	0	0.00	1	0.01
76 - Language Barrier - Other	3	0.20	11	0.57	44	3.09	58	2.45
77 - Refusal	70	5.67	296	18.87	425	24.51	791	21.84
78 - Parental Refusal	145	9.83	0	0.00	0	0.00	145	0.99
Other	9	0.65	26	1.83	24	1.27	59	1.28

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Colorado) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	376	100.00	390	100.00	422	100.00	1,188	100.00
70 - Interview Complete	319	84.84	301	77.18	307	72.75	927	78.03
71 - No One at DU	0	0.00	2	0.51	0	0.00	2	0.17
72 - Respondent Unavailable	6	1.60	14	3.59	22	5.21	42	3.54
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.27	5	1.28	2	0.47	8	0.67
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.26	4	0.95	5	0.42
77 - Refusal	10	2.66	58	14.87	86	20.38	154	12.96
78 - Parental Refusal	36	9.57	0	0.00	0	0.00	36	3.03
Other	4	1.06	9	2.31	1	0.24	14	1.18

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Colorado) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	376	100.00	390	100.00	422	100.00	1,188	100.00
70 - Interview Complete	319	86.13	301	78.11	307	73.13	927	74.95
71 - No One at DU	0	0.00	2	0.36	0	0.00	2	0.05
72 - Respondent Unavailable	6	1.63	14	3.44	22	4.72	42	4.27
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.23	5	0.86	2	0.64	8	0.63
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.15	4	1.15	5	0.92
77 - Refusal	10	2.73	58	14.82	86	20.25	154	17.96
78 - Parental Refusal	36	8.31	0	0.00	0	0.00	36	0.74
Other	4	0.98	9	2.26	1	0.11	14	0.47

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Connecticut) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	361	100.00	426	100.00	474	100.00	1,261	100.00
70 - Interview Complete	288	79.78	339	79.58	337	71.10	964	76.45
71 - No One at DU	1	0.28	7	1.64	8	1.69	16	1.27
72 - Respondent Unavailable	1	0.28	7	1.64	12	2.53	20	1.59
73 - Break Off (Partial Interview)	0	0.00	1	0.23	0	0.00	1	0.08
74 - Physically/Mentally Incompetent	1	0.28	1	0.23	13	2.74	15	1.19
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.28	0	0.00	3	0.63	4	0.32
77 - Refusal	13	3.60	65	15.26	98	20.68	176	13.96
78 - Parental Refusal	51	14.13	0	0.00	0	0.00	51	4.04
Other	5	1.39	6	1.41	3	0.63	14	1.11

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Connecticut) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	361	100.00	426	100.00	474	100.00	1,261	100.00
70 - Interview Complete	288	79.74	339	80.56	337	70.39	964	72.36
71 - No One at DU	1	0.19	7	1.55	8	0.98	16	0.97
72 - Respondent Unavailable	1	0.23	7	1.90	12	2.02	20	1.85
73 - Break Off (Partial Interview)	0	0.00	1	0.20	0	0.00	1	0.02
74 - Physically/Mentally Incompetent	1	0.34	1	0.19	13	3.48	15	2.83
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.23	0	0.00	3	1.23	4	1.00
77 - Refusal	13	3.83	65	13.96	98	21.35	176	19.01
78 - Parental Refusal	51	14.25	0	0.00	0	0.00	51	1.22
Other	5	1.18	6	1.64	3	0.55	14	0.73

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Delaware) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	376	100.00	305	100.00	429	100.00	1,110	100.00
70 - Interview Complete	307	81.65	246	80.66	340	79.25	893	80.45
71 - No One at DU	0	0.00	1	0.33	2	0.47	3	0.27
72 - Respondent Unavailable	0	0.00	2	0.66	2	0.47	4	0.36
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.53	4	1.31	4	0.93	10	0.90
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.23	1	0.09
77 - Refusal	10	2.66	48	15.74	80	18.65	138	12.43
78 - Parental Refusal	55	14.63	0	0.00	0	0.00	55	4.95
Other	2	0.53	4	1.31	0	0.00	6	0.54

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Delaware) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	376	100.00	305	100.00	429	100.00	1,110	100.00
70 - Interview Complete	307	82.59	246	83.85	340	79.02	893	79.90
71 - No One at DU	0	0.00	1	0.20	2	0.50	3	0.42
72 - Respondent Unavailable	0	0.00	2	0.65	2	0.48	4	0.46
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.50	4	0.66	4	0.99	10	0.91
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.15	1	0.12
77 - Refusal	10	2.39	48	13.09	80	18.85	138	16.81
78 - Parental Refusal	55	13.97	0	0.00	0	0.00	55	1.14
Other	2	0.55	4	1.56	0	0.00	6	0.23

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (District of Columbia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	362	100.00	398	100.00	365	100.00	1,125	100.00
70 - Interview Complete	329	90.88	344	86.43	289	79.18	962	85.51
71 - No One at DU	0	0.00	6	1.51	2	0.55	8	0.71
72 - Respondent Unavailable	3	0.83	7	1.76	7	1.92	17	1.51
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.55	0	0.00	8	2.19	10	0.89
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.28	0	0.00	4	1.10	5	0.44
77 - Refusal	7	1.93	40	10.05	53	14.52	100	8.89
78 - Parental Refusal	18	4.97	0	0.00	0	0.00	18	1.60
Other	2	0.55	1	0.25	2	0.55	5	0.44

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (District of Columbia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	362	100.00	398	100.00	365	100.00	1,125	100.00
70 - Interview Complete	329	91.77	344	87.06	289	78.39	962	80.64
71 - No One at DU	0	0.00	6	1.32	2	0.69	8	0.76
72 - Respondent Unavailable	3	0.73	7	2.09	7	2.03	17	1.97
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.48	0	0.00	8	2.40	10	1.87
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.30	0	0.00	4	1.01	5	0.80
77 - Refusal	7	1.91	40	9.23	53	14.82	100	13.12
78 - Parental Refusal	18	4.40	0	0.00	0	0.00	18	0.25
Other	2	0.41	1	0.31	2	0.67	5	0.59

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Florida) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,419	100.00	1,535	100.00	1,625	100.00	4,579	100.00
70 - Interview Complete	1,193	84.07	1,222	79.61	1,129	69.48	3,544	77.40
71 - No One at DU	9	0.63	9	0.59	7	0.43	25	0.55
72 - Respondent Unavailable	16	1.13	24	1.56	38	2.34	78	1.70
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.18	3	0.07
74 - Physically/Mentally Incompetent	17	1.20	14	0.91	33	2.03	64	1.40
75 - Language Barrier - Spanish	0	0.00	1	0.07	0	0.00	1	0.02
76 - Language Barrier - Other	2	0.14	3	0.20	17	1.05	22	0.48
77 - Refusal	29	2.04	217	14.14	359	22.09	605	13.21
78 - Parental Refusal	131	9.23	0	0.00	0	0.00	131	2.86
Other	22	1.55	45	2.93	39	2.40	106	2.31

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Florida) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,419	100.00	1,535	100.00	1,625	100.00	4,579	100.00
70 - Interview Complete	1,193	83.48	1,222	79.16	1,129	67.81	3,544	70.57
71 - No One at DU	9	0.63	9	0.51	7	0.43	25	0.45
72 - Respondent Unavailable	16	1.36	24	1.51	38	2.08	78	1.95
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.19	3	0.15
74 - Physically/Mentally Incompetent	17	1.20	14	1.43	33	2.76	64	2.46
75 - Language Barrier - Spanish	0	0.00	1	0.04	0	0.00	1	0.00
76 - Language Barrier - Other	2	0.13	3	0.17	17	1.29	22	1.06
77 - Refusal	29	1.91	217	14.30	359	23.39	605	20.40
78 - Parental Refusal	131	9.68	0	0.00	0	0.00	131	0.84
Other	22	1.61	45	2.89	39	2.06	106	2.12

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Georgia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	344	100.00	360	100.00	440	100.00	1,144	100.00
70 - Interview Complete	287	83.43	284	78.89	314	71.36	885	77.36
71 - No One at DU	0	0.00	0	0.00	2	0.45	2	0.17
72 - Respondent Unavailable	1	0.29	9	2.50	8	1.82	18	1.57
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.23	1	0.09
74 - Physically/Mentally Incompetent	9	2.62	2	0.56	9	2.05	20	1.75
75 - Language Barrier - Spanish	0	0.00	2	0.56	5	1.14	7	0.61
76 - Language Barrier - Other	0	0.00	0	0.00	6	1.36	6	0.52
77 - Refusal	22	6.40	53	14.72	86	19.55	161	14.07
78 - Parental Refusal	17	4.94	0	0.00	0	0.00	17	1.49
Other	8	2.33	10	2.78	9	2.05	27	2.36

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Georgia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	344	100.00	360	100.00	440	100.00	1,144	100.00
70 - Interview Complete	287	81.72	284	79.58	314	70.82	885	73.07
71 - No One at DU	0	0.00	0	0.00	2	0.68	2	0.52
72 - Respondent Unavailable	1	0.17	9	2.44	8	1.66	18	1.62
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.16	1	0.12
74 - Physically/Mentally Incompetent	9	3.47	2	0.66	9	2.07	20	2.02
75 - Language Barrier - Spanish	0	0.00	2	0.55	5	0.71	7	0.62
76 - Language Barrier - Other	0	0.00	0	0.00	6	1.55	6	1.19
77 - Refusal	22	5.82	53	13.56	86	19.68	161	17.48
78 - Parental Refusal	17	6.49	0	0.00	0	0.00	17	0.65
Other	8	2.34	10	3.21	9	2.67	27	2.71

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Hawaii) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	377	100.00	382	100.00	526	100.00	1,285	100.00
70 - Interview Complete	284	75.33	308	80.63	346	65.78	938	73.00
71 - No One at DU	8	2.12	9	2.36	14	2.66	31	2.41
72 - Respondent Unavailable	3	0.80	5	1.31	7	1.33	15	1.17
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.33	2	0.52	11	2.09	18	1.40
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.19	1	0.08
76 - Language Barrier - Other	4	1.06	1	0.26	18	3.42	23	1.79
77 - Refusal	11	2.92	52	13.61	124	23.57	187	14.55
78 - Parental Refusal	61	16.18	0	0.00	0	0.00	61	4.75
Other	1	0.27	5	1.31	5	0.95	11	0.86

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Hawaii) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	377	100.00	382	100.00	526	100.00	1,285	100.00
70 - Interview Complete	284	75.93	308	80.83	346	66.50	938	68.98
71 - No One at DU	8	1.50	9	2.47	14	2.13	31	2.12
72 - Respondent Unavailable	3	0.48	5	1.30	7	1.39	15	1.31
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.56	2	0.40	11	1.92	18	1.71
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.30	1	0.24
76 - Language Barrier - Other	4	0.91	1	0.17	18	3.82	23	3.15
77 - Refusal	11	3.27	52	13.60	124	23.03	187	20.30
78 - Parental Refusal	61	16.11	0	0.00	0	0.00	61	1.31
Other	1	0.23	5	1.23	5	0.91	11	0.89

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Idaho) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	389	100.00	334	100.00	413	100.00	1,136	100.00
70 - Interview Complete	345	88.69	262	78.44	314	76.03	921	81.07
71 - No One at DU	0	0.00	1	0.30	3	0.73	4	0.35
72 - Respondent Unavailable	2	0.51	3	0.90	1	0.24	6	0.53
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.03	1	0.30	1	0.24	6	0.53
75 - Language Barrier - Spanish	0	0.00	2	0.60	4	0.97	6	0.53
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.24	1	0.09
77 - Refusal	15	3.86	63	18.86	89	21.55	167	14.70
78 - Parental Refusal	20	5.14	0	0.00	0	0.00	20	1.76
Other	3	0.77	2	0.60	0	0.00	5	0.44

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Idaho) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	389	100.00	334	100.00	413	100.00	1,136	100.00
70 - Interview Complete	345	88.85	262	80.12	314	76.28	921	78.38
71 - No One at DU	0	0.00	1	0.24	3	0.64	4	0.50
72 - Respondent Unavailable	2	0.42	3	0.64	1	0.24	6	0.32
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	0.84	1	0.22	1	0.45	6	0.47
75 - Language Barrier - Spanish	0	0.00	2	0.42	4	0.90	6	0.72
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.57	1	0.41
77 - Refusal	15	3.99	63	17.89	89	20.92	167	18.41
78 - Parental Refusal	20	5.27	0	0.00	0	0.00	20	0.64
Other	3	0.63	2	0.48	0	0.00	5	0.15

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Illinois) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,517	100.00	1,562	100.00	1,792	100.00	4,871	100.00
70 - Interview Complete	1,234	81.34	1,190	76.18	1,248	69.64	3,672	75.38
71 - No One at DU	18	1.19	60	3.84	57	3.18	135	2.77
72 - Respondent Unavailable	15	0.99	30	1.92	32	1.79	77	1.58
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	15	0.99	19	1.22	32	1.79	66	1.35
75 - Language Barrier - Spanish	0	0.00	1	0.06	0	0.00	1	0.02
76 - Language Barrier - Other	2	0.13	3	0.19	35	1.95	40	0.82
77 - Refusal	71	4.68	240	15.36	379	21.15	690	14.17
78 - Parental Refusal	154	10.15	0	0.00	0	0.00	154	3.16
Other	8	0.53	19	1.22	9	0.50	36	0.74

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Illinois) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,517	100.00	1,562	100.00	1,792	100.00	4,871	100.00
70 - Interview Complete	1,234	81.95	1,190	76.45	1,248	68.62	3,672	70.95
71 - No One at DU	18	1.07	60	3.76	57	2.78	135	2.74
72 - Respondent Unavailable	15	1.02	30	1.88	32	1.92	77	1.83
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	15	0.84	19	1.28	32	2.51	66	2.19
75 - Language Barrier - Spanish	0	0.00	1	0.05	0	0.00	1	0.01
76 - Language Barrier - Other	2	0.27	3	0.33	35	2.15	40	1.73
77 - Refusal	71	4.49	240	14.68	379	21.30	690	18.78
78 - Parental Refusal	154	9.75	0	0.00	0	0.00	154	0.97
Other	8	0.60	19	1.57	9	0.72	36	0.82

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Indiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	408	100.00	433	100.00	1,171	100.00
70 - Interview Complete	271	82.12	328	80.39	312	72.06	911	77.80
71 - No One at DU	4	1.21	6	1.47	4	0.92	14	1.20
72 - Respondent Unavailable	1	0.30	7	1.72	8	1.85	16	1.37
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.52	6	1.47	8	1.85	19	1.62
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.25	0	0.00	1	0.09
77 - Refusal	5	1.52	60	14.71	99	22.86	164	14.01
78 - Parental Refusal	43	13.03	0	0.00	0	0.00	43	3.67
Other	1	0.30	0	0.00	2	0.46	3	0.26

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Indiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	408	100.00	433	100.00	1,171	100.00
70 - Interview Complete	271	82.24	328	80.64	312	70.63	911	72.95
71 - No One at DU	4	1.19	6	1.44	4	0.98	14	1.05
72 - Respondent Unavailable	1	0.30	7	1.88	8	2.08	16	1.89
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	2.14	6	1.24	8	2.03	19	1.94
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.46	0	0.00	1	0.06
77 - Refusal	5	1.51	60	14.33	99	23.70	164	20.49
78 - Parental Refusal	43	12.38	0	0.00	0	0.00	43	1.14
Other	1	0.25	0	0.00	2	0.57	3	0.47

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Iowa) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	373	100.00	362	100.00	402	100.00	1,137	100.00
70 - Interview Complete	314	84.18	287	79.28	299	74.38	900	79.16
71 - No One at DU	1	0.27	8	2.21	0	0.00	9	0.79
72 - Respondent Unavailable	2	0.54	9	2.49	6	1.49	17	1.50
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.34	4	1.10	4	1.00	13	1.14
75 - Language Barrier - Spanish	0	0.00	1	0.28	0	0.00	1	0.09
76 - Language Barrier - Other	0	0.00	2	0.55	1	0.25	3	0.26
77 - Refusal	8	2.14	42	11.60	85	21.14	135	11.87
78 - Parental Refusal	42	11.26	0	0.00	0	0.00	42	3.69
Other	1	0.27	9	2.49	7	1.74	17	1.50

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Iowa) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	373	100.00	362	100.00	402	100.00	1,137	100.00
70 - Interview Complete	314	82.15	287	79.41	299	72.90	900	74.74
71 - No One at DU	1	0.17	8	2.21	0	0.00	9	0.32
72 - Respondent Unavailable	2	0.49	9	2.47	6	2.00	17	1.91
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	3.89	4	1.20	4	1.09	13	1.39
75 - Language Barrier - Spanish	0	0.00	1	0.20	0	0.00	1	0.03
76 - Language Barrier - Other	0	0.00	2	0.23	1	0.15	3	0.15
77 - Refusal	8	2.04	42	11.92	85	22.47	135	18.92
78 - Parental Refusal	42	11.05	0	0.00	0	0.00	42	1.13
Other	1	0.21	9	2.38	7	1.40	17	1.41

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Kansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	388	100.00	318	100.00	403	100.00	1,109	100.00
70 - Interview Complete	343	88.40	265	83.33	304	75.43	912	82.24
71 - No One at DU	1	0.26	0	0.00	6	1.49	7	0.63
72 - Respondent Unavailable	3	0.77	5	1.57	6	1.49	14	1.26
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.52	4	1.26	4	0.99	10	0.90
75 - Language Barrier - Spanish	0	0.00	1	0.31	1	0.25	2	0.18
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.99	4	0.36
77 - Refusal	13	3.35	34	10.69	77	19.11	124	11.18
78 - Parental Refusal	24	6.19	0	0.00	0	0.00	24	2.16
Other	2	0.52	9	2.83	1	0.25	12	1.08

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Kansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	388	100.00	318	100.00	403	100.00	1,109	100.00
70 - Interview Complete	343	88.15	265	84.49	304	75.30	912	77.88
71 - No One at DU	1	0.28	0	0.00	6	1.43	7	1.12
72 - Respondent Unavailable	3	0.72	5	1.55	6	1.38	14	1.33
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.57	4	1.15	4	1.16	10	1.10
75 - Language Barrier - Spanish	0	0.00	1	0.26	1	0.39	2	0.33
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.94	4	0.72
77 - Refusal	13	3.17	34	9.48	77	19.27	124	16.26
78 - Parental Refusal	24	6.63	0	0.00	0	0.00	24	0.70
Other	2	0.48	9	3.07	1	0.14	12	0.57

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Kentucky) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	384	100.00	380	100.00	420	100.00	1,184	100.00
70 - Interview Complete	318	82.81	302	79.47	307	73.10	927	78.29
71 - No One at DU	3	0.78	8	2.11	6	1.43	17	1.44
72 - Respondent Unavailable	6	1.56	10	2.63	6	1.43	22	1.86
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.30	4	1.05	12	2.86	21	1.77
75 - Language Barrier - Spanish	0	0.00	2	0.53	2	0.48	4	0.34
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.24	1	0.08
77 - Refusal	13	3.39	43	11.32	82	19.52	138	11.66
78 - Parental Refusal	36	9.38	0	0.00	0	0.00	36	3.04
Other	3	0.78	11	2.89	4	0.95	18	1.52

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Kentucky) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	384	100.00	380	100.00	420	100.00	1,184	100.00
70 - Interview Complete	318	81.85	302	80.21	307	71.39	927	73.49
71 - No One at DU	3	0.66	8	1.86	6	1.27	17	1.28
72 - Respondent Unavailable	6	1.60	10	2.22	6	1.02	22	1.23
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.22	4	0.81	12	3.58	21	3.01
75 - Language Barrier - Spanish	0	0.00	2	0.32	2	0.19	4	0.19
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.19	1	0.15
77 - Refusal	13	3.70	43	10.12	82	21.40	138	18.31
78 - Parental Refusal	36	10.06	0	0.00	0	0.00	36	0.95
Other	3	0.91	11	4.46	4	0.95	18	1.39

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Louisiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	364	100.00	406	100.00	1,100	100.00
70 - Interview Complete	292	88.48	303	83.24	306	75.37	901	81.91
71 - No One at DU	2	0.61	11	3.02	5	1.23	18	1.64
72 - Respondent Unavailable	3	0.91	4	1.10	7	1.72	14	1.27
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	0.91	5	1.37	8	1.97	16	1.45
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.27	1	0.25	2	0.18
77 - Refusal	5	1.52	36	9.89	76	18.72	117	10.64
78 - Parental Refusal	24	7.27	0	0.00	0	0.00	24	2.18
Other	1	0.30	4	1.10	3	0.74	8	0.73

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Louisiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	364	100.00	406	100.00	1,100	100.00
70 - Interview Complete	292	88.75	303	82.65	306	75.23	901	77.61
71 - No One at DU	2	1.01	11	3.71	5	1.32	18	1.63
72 - Respondent Unavailable	3	0.53	4	1.02	7	1.17	14	1.08
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	0.83	5	1.29	8	2.51	16	2.17
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.39	1	0.60	2	0.51
77 - Refusal	5	1.53	36	9.57	76	18.45	117	15.52
78 - Parental Refusal	24	7.14	0	0.00	0	0.00	24	0.70
Other	1	0.22	4	1.38	3	0.73	8	0.77

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Maine) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	359	100.00	387	100.00	388	100.00	1,134	100.00
70 - Interview Complete	305	84.96	325	83.98	308	79.38	938	82.72
71 - No One at DU	4	1.11	5	1.29	4	1.03	13	1.15
72 - Respondent Unavailable	4	1.11	3	0.78	3	0.77	10	0.88
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.56	5	1.29	8	2.06	15	1.32
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	2	0.52	1	0.26	3	0.26
77 - Refusal	10	2.79	44	11.37	63	16.24	117	10.32
78 - Parental Refusal	34	9.47	0	0.00	0	0.00	34	3.00
Other	0	0.00	3	0.78	1	0.26	4	0.35

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Maine) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	359	100.00	387	100.00	388	100.00	1,134	100.00
70 - Interview Complete	305	85.30	325	84.13	308	77.79	938	79.20
71 - No One at DU	4	1.20	5	1.27	4	0.70	13	0.81
72 - Respondent Unavailable	4	0.96	3	0.83	3	1.03	10	1.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.85	5	1.58	8	2.36	15	2.14
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	2	0.61	1	0.92	3	0.80
77 - Refusal	10	2.73	44	10.39	63	17.00	117	14.94
78 - Parental Refusal	34	8.96	0	0.00	0	0.00	34	0.81
Other	0	0.00	3	1.19	1	0.18	4	0.28

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Maryland) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	363	100.00	381	100.00	1,074	100.00
70 - Interview Complete	282	85.45	306	84.30	286	75.07	874	81.38
71 - No One at DU	0	0.00	2	0.55	0	0.00	2	0.19
72 - Respondent Unavailable	3	0.91	6	1.65	1	0.26	10	0.93
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.30	2	0.55	11	2.89	14	1.30
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.28	6	1.57	7	0.65
77 - Refusal	8	2.42	44	12.12	76	19.95	128	11.92
78 - Parental Refusal	35	10.61	0	0.00	0	0.00	35	3.26
Other	1	0.30	2	0.55	1	0.26	4	0.37

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Maryland) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	330	100.00	363	100.00	381	100.00	1,074	100.00
70 - Interview Complete	282	85.48	306	83.31	286	73.39	874	75.90
71 - No One at DU	0	0.00	2	0.42	0	0.00	2	0.06
72 - Respondent Unavailable	3	0.71	6	1.77	1	0.19	10	0.45
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.24	2	0.36	11	3.77	14	2.97
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.22	6	1.61	7	1.26
77 - Refusal	8	2.13	44	13.42	76	20.86	128	18.03
78 - Parental Refusal	35	11.09	0	0.00	0	0.00	35	1.09
Other	1	0.35	2	0.50	1	0.18	4	0.24

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Massachusetts) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	380	100.00	408	100.00	465	100.00	1,253	100.00
70 - Interview Complete	309	81.32	312	76.47	334	71.83	955	76.22
71 - No One at DU	0	0.00	7	1.72	6	1.29	13	1.04
72 - Respondent Unavailable	5	1.32	21	5.15	11	2.37	37	2.95
73 - Break Off (Partial Interview)	0	0.00	1	0.25	1	0.22	2	0.16
74 - Physically/Mentally Incompetent	4	1.05	3	0.74	9	1.94	16	1.28
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.22	1	0.08
76 - Language Barrier - Other	0	0.00	4	0.98	13	2.80	17	1.36
77 - Refusal	15	3.95	49	12.01	83	17.85	147	11.73
78 - Parental Refusal	45	11.84	0	0.00	0	0.00	45	3.59
Other	2	0.53	11	2.70	7	1.51	20	1.60

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Massachusetts) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	380	100.00	408	100.00	465	100.00	1,253	100.00
70 - Interview Complete	309	81.19	312	77.20	334	69.50	955	71.52
71 - No One at DU	0	0.00	7	1.64	6	1.01	13	1.01
72 - Respondent Unavailable	5	0.80	21	4.53	11	1.96	37	2.20
73 - Break Off (Partial Interview)	0	0.00	1	0.16	1	0.30	2	0.25
74 - Physically/Mentally Incompetent	4	1.06	3	0.66	9	3.36	16	2.80
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.09	1	0.07
76 - Language Barrier - Other	0	0.00	4	0.73	13	3.40	17	2.76
77 - Refusal	15	5.18	49	12.59	83	19.08	147	17.04
78 - Parental Refusal	45	11.54	0	0.00	0	0.00	45	0.98
Other	2	0.23	11	2.48	7	1.30	20	1.37

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Michigan) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,445	100.00	1,508	100.00	1,653	100.00	4,606	100.00
70 - Interview Complete	1,178	81.52	1,231	81.63	1,246	75.38	3,655	79.35
71 - No One at DU	8	0.55	16	1.06	16	0.97	40	0.87
72 - Respondent Unavailable	13	0.90	25	1.66	14	0.85	52	1.13
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	9	0.62	12	0.80	24	1.45	45	0.98
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.07	8	0.53	15	0.91	24	0.52
77 - Refusal	46	3.18	196	13.00	334	20.21	576	12.51
78 - Parental Refusal	178	12.32	0	0.00	0	0.00	178	3.86
Other	12	0.83	20	1.33	4	0.24	36	0.78

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Michigan) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,445	100.00	1,508	100.00	1,653	100.00	4,606	100.00
70 - Interview Complete	1,178	81.72	1,231	81.78	1,246	73.97	3,655	75.75
71 - No One at DU	8	0.49	16	0.89	16	0.78	40	0.77
72 - Respondent Unavailable	13	0.92	25	1.40	14	0.72	52	0.83
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	9	0.57	12	0.82	24	1.92	45	1.64
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.07	8	0.55	15	1.33	24	1.11
77 - Refusal	46	3.05	196	13.23	334	21.00	576	18.23
78 - Parental Refusal	178	12.33	0	0.00	0	0.00	178	1.20
Other	12	0.86	20	1.33	4	0.28	36	0.47

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Minnesota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	363	100.00	339	100.00	390	100.00	1,092	100.00
70 - Interview Complete	324	89.26	272	80.24	306	78.46	902	82.60
71 - No One at DU	0	0.00	5	1.47	4	1.03	9	0.82
72 - Respondent Unavailable	2	0.55	4	1.18	2	0.51	8	0.73
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.28	0	0.00	0	0.00	1	0.09
75 - Language Barrier - Spanish	0	0.00	1	0.29	4	1.03	5	0.46
76 - Language Barrier - Other	1	0.28	1	0.29	8	2.05	10	0.92
77 - Refusal	5	1.38	48	14.16	62	15.90	115	10.53
78 - Parental Refusal	30	8.26	0	0.00	0	0.00	30	2.75
Other	0	0.00	8	2.36	4	1.03	12	1.10

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Minnesota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	363	100.00	339	100.00	390	100.00	1,092	100.00
70 - Interview Complete	324	89.54	272	79.91	306	80.32	902	81.16
71 - No One at DU	0	0.00	5	1.20	4	0.63	9	0.65
72 - Respondent Unavailable	2	0.38	4	0.84	2	0.39	8	0.45
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.20	0	0.00	0	0.00	1	0.02
75 - Language Barrier - Spanish	0	0.00	1	0.20	4	0.40	5	0.34
76 - Language Barrier - Other	1	0.48	1	0.44	8	2.86	10	2.31
77 - Refusal	5	1.28	48	15.55	62	14.87	115	13.64
78 - Parental Refusal	30	8.11	0	0.00	0	0.00	30	0.79
Other	0	0.00	8	1.86	4	0.52	12	0.65

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Mississippi) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	384	100.00	338	100.00	378	100.00	1,100	100.00
70 - Interview Complete	313	81.51	297	87.87	291	76.98	901	81.91
71 - No One at DU	7	1.82	12	3.55	10	2.65	29	2.64
72 - Respondent Unavailable	2	0.52	4	1.18	9	2.38	15	1.36
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.04	1	0.30	10	2.65	15	1.36
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.26	1	0.09
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	2.86	15	4.44	52	13.76	78	7.09
78 - Parental Refusal	37	9.64	0	0.00	0	0.00	37	3.36
Other	10	2.60	9	2.66	5	1.32	24	2.18

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Mississippi) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	384	100.00	338	100.00	378	100.00	1,100	100.00
70 - Interview Complete	313	80.62	297	88.22	291	76.36	901	78.58
71 - No One at DU	7	2.02	12	3.61	10	2.20	29	2.39
72 - Respondent Unavailable	2	0.62	4	1.42	9	2.31	15	2.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.05	1	0.07	10	3.43	15	2.67
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.16	1	0.12
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	3.69	15	4.18	52	14.34	78	11.67
78 - Parental Refusal	37	8.91	0	0.00	0	0.00	37	0.97
Other	10	3.09	9	2.51	5	1.19	24	1.59

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Missouri) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	367	100.00	356	100.00	426	100.00	1,149	100.00
70 - Interview Complete	312	85.01	290	81.46	313	73.47	915	79.63
71 - No One at DU	5	1.36	6	1.69	7	1.64	18	1.57
72 - Respondent Unavailable	5	1.36	2	0.56	6	1.41	13	1.13
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	1.91	6	1.69	8	1.88	21	1.83
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	5	1.36	48	13.48	92	21.60	145	12.62
78 - Parental Refusal	32	8.72	0	0.00	0	0.00	32	2.79
Other	1	0.27	4	1.12	0	0.00	5	0.44

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Missouri) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	367	100.00	356	100.00	426	100.00	1,149	100.00
70 - Interview Complete	312	85.89	290	82.34	313	71.53	915	74.36
71 - No One at DU	5	1.03	6	1.30	7	1.82	18	1.68
72 - Respondent Unavailable	5	1.30	2	0.41	6	1.22	13	1.12
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	2.25	6	1.84	8	2.62	21	2.48
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	5	1.40	48	12.58	92	22.80	145	19.35
78 - Parental Refusal	32	7.69	0	0.00	0	0.00	32	0.76
Other	1	0.45	4	1.54	0	0.00	5	0.25

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Montana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	388	100.00	350	100.00	371	100.00	1,109	100.00
70 - Interview Complete	316	81.44	279	79.71	281	75.74	876	78.99
71 - No One at DU	0	0.00	1	0.29	3	0.81	4	0.36
72 - Respondent Unavailable	1	0.26	4	1.14	2	0.54	7	0.63
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.03	1	0.29	3	0.81	8	0.72
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	2.84	59	16.86	81	21.83	151	13.62
78 - Parental Refusal	56	14.43	0	0.00	0	0.00	56	5.05
Other	0	0.00	6	1.71	1	0.27	7	0.63

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Montana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	388	100.00	350	100.00	371	100.00	1,109	100.00
70 - Interview Complete	316	81.81	279	78.48	281	76.71	876	77.46
71 - No One at DU	0	0.00	1	0.54	3	0.71	4	0.62
72 - Respondent Unavailable	1	0.25	4	1.16	2	0.64	7	0.67
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.30	1	0.42	3	1.11	8	1.03
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	2.59	59	17.63	81	20.65	151	18.44
78 - Parental Refusal	56	14.05	0	0.00	0	0.00	56	1.39
Other	0	0.00	6	1.78	1	0.19	7	0.39

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Nebraska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	322	100.00	433	100.00	415	100.00	1,170	100.00
70 - Interview Complete	278	86.34	365	84.30	297	71.57	940	80.34
71 - No One at DU	2	0.62	10	2.31	5	1.20	17	1.45
72 - Respondent Unavailable	5	1.55	16	3.70	8	1.93	29	2.48
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.62	1	0.23	6	1.45	9	0.77
75 - Language Barrier - Spanish	0	0.00	1	0.23	0	0.00	1	0.09
76 - Language Barrier - Other	0	0.00	5	1.15	1	0.24	6	0.51
77 - Refusal	5	1.55	34	7.85	96	23.13	135	11.54
78 - Parental Refusal	30	9.32	0	0.00	0	0.00	30	2.56
Other	0	0.00	1	0.23	2	0.48	3	0.26

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Nebraska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	322	100.00	433	100.00	415	100.00	1,170	100.00
70 - Interview Complete	278	86.79	365	84.84	297	69.50	940	73.14
71 - No One at DU	2	0.51	10	2.08	5	1.46	17	1.45
72 - Respondent Unavailable	5	1.52	16	4.05	8	2.46	29	2.57
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.52	1	0.16	6	1.82	9	1.48
75 - Language Barrier - Spanish	0	0.00	1	0.40	0	0.00	1	0.05
76 - Language Barrier - Other	0	0.00	5	1.25	1	0.25	6	0.35
77 - Refusal	5	1.33	34	6.89	96	23.98	135	19.60
78 - Parental Refusal	30	9.33	0	0.00	0	0.00	30	0.90
Other	0	0.00	1	0.33	2	0.53	3	0.45

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Nevada) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	333	100.00	368	100.00	433	100.00	1,134	100.00
70 - Interview Complete	290	87.09	289	78.53	324	74.83	903	79.63
71 - No One at DU	2	0.60	4	1.09	5	1.15	11	0.97
72 - Respondent Unavailable	1	0.30	2	0.54	4	0.92	7	0.62
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.30	2	0.54	5	1.15	8	0.71
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.23	1	0.09
76 - Language Barrier - Other	0	0.00	1	0.27	2	0.46	3	0.26
77 - Refusal	15	4.50	63	17.12	89	20.55	167	14.73
78 - Parental Refusal	22	6.61	0	0.00	0	0.00	22	1.94
Other	2	0.60	7	1.90	3	0.69	12	1.06

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Nevada) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	333	100.00	368	100.00	433	100.00	1,134	100.00
70 - Interview Complete	290	86.58	289	79.10	324	73.75	903	75.62
71 - No One at DU	2	0.90	4	1.10	5	1.32	11	1.25
72 - Respondent Unavailable	1	0.50	2	0.69	4	0.67	7	0.65
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.26	2	0.67	5	1.00	8	0.89
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.13	1	0.10
76 - Language Barrier - Other	0	0.00	1	0.21	2	0.40	3	0.34
77 - Refusal	15	5.39	63	16.46	89	22.32	167	20.00
78 - Parental Refusal	22	6.06	0	0.00	0	0.00	22	0.58
Other	2	0.32	7	1.77	3	0.41	12	0.57

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (New Hampshire) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	405	100.00	417	100.00	437	100.00	1,259	100.00
70 - Interview Complete	305	75.31	324	77.70	321	73.46	950	75.46
71 - No One at DU	1	0.25	3	0.72	1	0.23	5	0.40
72 - Respondent Unavailable	1	0.25	9	2.16	2	0.46	12	0.95
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	10	2.47	3	0.72	6	1.37	19	1.51
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.25	0	0.00	5	1.14	6	0.48
77 - Refusal	10	2.47	75	17.99	102	23.34	187	14.85
78 - Parental Refusal	77	19.01	0	0.00	0	0.00	77	6.12
Other	0	0.00	3	0.72	0	0.00	3	0.24

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (New Hampshire) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	405	100.00	417	100.00	437	100.00	1,259	100.00
70 - Interview Complete	305	75.51	324	78.95	321	71.84	950	73.08
71 - No One at DU	1	0.21	3	0.61	1	0.20	5	0.25
72 - Respondent Unavailable	1	0.32	9	1.71	2	0.32	12	0.49
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	10	2.05	3	0.58	6	1.56	19	1.48
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.31	0	0.00	5	1.39	6	1.12
77 - Refusal	10	2.50	75	17.41	102	24.69	187	21.69
78 - Parental Refusal	77	19.11	0	0.00	0	0.00	77	1.79
Other	0	0.00	3	0.74	0	0.00	3	0.09

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (New Jersey) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	349	100.00	378	100.00	428	100.00	1,155	100.00
70 - Interview Complete	291	83.38	292	77.25	315	73.60	898	77.75
71 - No One at DU	0	0.00	5	1.32	2	0.47	7	0.61
72 - Respondent Unavailable	0	0.00	2	0.53	1	0.23	3	0.26
73 - Break Off (Partial Interview)	0	0.00	2	0.53	1	0.23	3	0.26
74 - Physically/Mentally Incompetent	3	0.86	3	0.79	8	1.87	14	1.21
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.23	1	0.09
76 - Language Barrier - Other	0	0.00	0	0.00	5	1.17	5	0.43
77 - Refusal	7	2.01	69	18.25	92	21.50	168	14.55
78 - Parental Refusal	48	13.75	0	0.00	0	0.00	48	4.16
Other	0	0.00	5	1.32	3	0.70	8	0.69

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (New Jersey) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	349	100.00	378	100.00	428	100.00	1,155	100.00
70 - Interview Complete	291	83.09	292	78.25	315	71.73	898	73.64
71 - No One at DU	0	0.00	5	1.27	2	0.44	7	0.50
72 - Respondent Unavailable	0	0.00	2	0.38	1	0.10	3	0.13
73 - Break Off (Partial Interview)	0	0.00	2	0.41	1	0.14	3	0.16
74 - Physically/Mentally Incompetent	3	1.14	3	0.92	8	2.56	14	2.22
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.09	1	0.07
76 - Language Barrier - Other	0	0.00	0	0.00	5	1.76	5	1.37
77 - Refusal	7	1.76	69	17.56	92	22.79	168	20.11
78 - Parental Refusal	48	14.01	0	0.00	0	0.00	48	1.35
Other	0	0.00	5	1.23	3	0.37	8	0.44

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (New Mexico) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	332	100.00	369	100.00	400	100.00	1,101	100.00
70 - Interview Complete	290	87.35	303	82.11	286	71.50	879	79.84
71 - No One at DU	1	0.30	3	0.81	5	1.25	9	0.82
72 - Respondent Unavailable	0	0.00	2	0.54	2	0.50	4	0.36
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.60	2	0.54	5	1.25	9	0.82
75 - Language Barrier - Spanish	0	0.00	1	0.27	0	0.00	1	0.09
76 - Language Barrier - Other	0	0.00	1	0.27	4	1.00	5	0.45
77 - Refusal	12	3.61	53	14.36	94	23.50	159	14.44
78 - Parental Refusal	25	7.53	0	0.00	0	0.00	25	2.27
Other	2	0.60	4	1.08	4	1.00	10	0.91

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (New Mexico) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	332	100.00	369	100.00	400	100.00	1,101	100.00
70 - Interview Complete	290	87.22	303	81.39	286	71.17	879	74.17
71 - No One at DU	1	0.21	3	0.55	5	1.17	9	0.99
72 - Respondent Unavailable	0	0.00	2	0.50	2	0.27	4	0.27
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.43	2	0.33	5	1.81	9	1.48
75 - Language Barrier - Spanish	0	0.00	1	0.23	0	0.00	1	0.03
76 - Language Barrier - Other	0	0.00	1	0.58	4	2.29	5	1.83
77 - Refusal	12	3.46	53	15.40	94	21.97	159	19.20
78 - Parental Refusal	25	8.14	0	0.00	0	0.00	25	0.84
Other	2	0.54	4	1.02	4	1.31	10	1.19

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (New York) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,564	100.00	1,778	100.00	1,925	100.00	5,267	100.00
70 - Interview Complete	1,193	76.28	1,266	71.20	1,221	63.43	3,680	69.87
71 - No One at DU	13	0.83	50	2.81	59	3.06	122	2.32
72 - Respondent Unavailable	33	2.11	87	4.89	84	4.36	204	3.87
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.05	1	0.02
74 - Physically/Mentally Incompetent	22	1.41	10	0.56	30	1.56	62	1.18
75 - Language Barrier - Spanish	0	0.00	1	0.06	0	0.00	1	0.02
76 - Language Barrier - Other	5	0.32	9	0.51	41	2.13	55	1.04
77 - Refusal	50	3.20	319	17.94	470	24.42	839	15.93
78 - Parental Refusal	242	15.47	0	0.00	0	0.00	242	4.59
Other	6	0.38	36	2.02	19	0.99	61	1.16

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (New York) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,564	100.00	1,778	100.00	1,925	100.00	5,267	100.00
70 - Interview Complete	1,193	75.84	1,266	71.75	1,221	61.76	3,680	64.38
71 - No One at DU	13	0.78	50	2.49	59	2.52	122	2.36
72 - Respondent Unavailable	33	2.11	87	4.74	84	4.52	204	4.33
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.09	1	0.07
74 - Physically/Mentally Incompetent	22	1.58	10	0.43	30	1.99	62	1.74
75 - Language Barrier - Spanish	0	0.00	1	0.04	0	0.00	1	0.01
76 - Language Barrier - Other	5	0.34	9	0.61	41	3.22	55	2.60
77 - Refusal	50	3.34	319	17.62	470	24.84	839	21.94
78 - Parental Refusal	242	15.54	0	0.00	0	0.00	242	1.39
Other	6	0.48	36	2.31	19	1.06	61	1.18

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (North Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	354	100.00	382	100.00	381	100.00	1,117	100.00
70 - Interview Complete	298	84.18	337	88.22	282	74.02	917	82.09
71 - No One at DU	4	1.13	2	0.52	2	0.52	8	0.72
72 - Respondent Unavailable	4	1.13	2	0.52	8	2.10	14	1.25
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.69	2	0.52	12	3.15	20	1.79
75 - Language Barrier - Spanish	0	0.00	1	0.26	2	0.52	3	0.27
76 - Language Barrier - Other	1	0.28	1	0.26	0	0.00	2	0.18
77 - Refusal	3	0.85	35	9.16	72	18.90	110	9.85
78 - Parental Refusal	36	10.17	0	0.00	0	0.00	36	3.22
Other	2	0.56	2	0.52	3	0.79	7	0.63

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (North Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	354	100.00	382	100.00	381	100.00	1,117	100.00
70 - Interview Complete	298	83.53	337	87.89	282	72.36	917	75.46
71 – No One at DU	4	1.15	2	0.39	2	0.83	8	0.81
72 - Respondent Unavailable	4	1.13	2	0.58	8	1.46	14	1.31
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.61	2	0.40	12	4.20	20	3.46
75 - Language Barrier - Spanish	0	0.00	1	0.35	2	0.47	3	0.41
76 - Language Barrier - Other	1	0.34	1	0.14	0	0.00	2	0.05
77 - Refusal	3	1.00	35	9.86	72	19.92	110	16.74
78 - Parental Refusal	36	10.52	0	0.00	0	0.00	36	1.06
Other	2	0.71	2	0.37	3	0.75	7	0.70

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (North Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	371	100.00	339	100.00	446	100.00	1,156	100.00
70 - Interview Complete	309	83.29	268	79.06	318	71.30	895	77.42
71 - No One at DU	0	0.00	0	0.00	1	0.22	1	0.09
72 - Respondent Unavailable	1	0.27	5	1.47	6	1.35	12	1.04
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.27	3	0.88	2	0.45	6	0.52
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.22	1	0.09
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.22	1	0.09
77 - Refusal	27	7.28	61	17.99	116	26.01	204	17.65
78 - Parental Refusal	33	8.89	0	0.00	0	0.00	33	2.85
Other	0	0.00	2	0.59	1	0.22	3	0.26

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (North Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	371	100.00	339	100.00	446	100.00	1,156	100.00
70 - Interview Complete	309	83.61	268	79.86	318	70.99	895	73.47
71 - No One at DU	0	0.00	0	0.00	1	0.10	1	0.07
72 - Respondent Unavailable	1	0.19	5	1.28	6	1.06	12	1.02
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.22	3	0.64	2	0.32	6	0.36
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.14	1	0.11
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.39	1	0.29
77 - Refusal	27	6.59	61	17.70	116	26.86	204	23.68
78 - Parental Refusal	33	9.40	0	0.00	0	0.00	33	0.80
Other	0	0.00	2	0.52	1	0.15	3	0.19

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Ohio) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,628	100.00	1,475	100.00	1,724	100.00	4,827	100.00
70 - Interview Complete	1,297	79.67	1,148	77.83	1,242	72.04	3,687	76.38
71 - No One at DU	18	1.11	31	2.10	23	1.33	72	1.49
72 - Respondent Unavailable	20	1.23	30	2.03	31	1.80	81	1.68
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	20	1.23	12	0.81	35	2.03	67	1.39
75 - Language Barrier - Spanish	1	0.06	4	0.27	3	0.17	8	0.17
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.23	4	0.08
77 - Refusal	59	3.62	221	14.98	376	21.81	656	13.59
78 - Parental Refusal	199	12.22	0	0.00	0	0.00	199	4.12
Other	14	0.86	29	1.97	10	0.58	53	1.10

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Ohio) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,628	100.00	1,475	100.00	1,724	100.00	4,827	100.00
70 - Interview Complete	1,297	79.72	1,148	77.78	1,242	71.02	3,687	72.73
71 - No One at DU	18	1.30	31	2.02	23	1.45	72	1.50
72 - Respondent Unavailable	20	1.26	30	2.08	31	1.76	81	1.75
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	20	1.46	12	0.78	35	2.76	67	2.39
75 - Language Barrier - Spanish	1	0.06	4	0.27	3	0.14	8	0.15
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.20	4	0.15
77 - Refusal	59	3.48	221	15.02	376	22.12	656	19.38
78 - Parental Refusal	199	11.88	0	0.00	0	0.00	199	1.18
Other	14	0.85	29	2.04	10	0.55	53	0.77

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Oklahoma) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	385	100.00	383	100.00	421	100.00	1,189	100.00
70 - Interview Complete	303	78.70	297	77.55	308	73.16	908	76.37
71 - No One at DU	1	0.26	2	0.52	1	0.24	4	0.34
72 - Respondent Unavailable	4	1.04	2	0.52	5	1.19	11	0.93
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.30	2	0.52	8	1.90	15	1.26
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	13	3.38	71	18.54	96	22.80	180	15.14
78 - Parental Refusal	58	15.06	0	0.00	0	0.00	58	4.88
Other	1	0.26	9	2.35	3	0.71	13	1.09

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Oklahoma) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	385	100.00	383	100.00	421	100.00	1,189	100.00
70 - Interview Complete	303	78.05	297	76.87	308	70.82	908	72.38
71 - No One at DU	1	0.19	2	0.41	1	0.31	4	0.31
72 - Respondent Unavailable	4	1.32	2	0.60	5	1.28	11	1.18
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.09	2	0.75	8	2.05	15	1.77
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	13	3.54	71	19.29	96	24.94	180	22.03
78 - Parental Refusal	58	15.59	0	0.00	0	0.00	58	1.54
Other	1	0.24	9	2.08	3	0.61	13	0.78

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Oregon) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	311	100.00	407	100.00	447	100.00	1,165	100.00
70 - Interview Complete	270	86.82	318	78.13	335	74.94	923	79.23
71 - No One at DU	0	0.00	5	1.23	3	0.67	8	0.69
72 - Respondent Unavailable	5	1.61	8	1.97	14	3.13	27	2.32
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.29	2	0.49	5	1.12	11	0.94
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	0.45	2	0.17
76 - Language Barrier - Other	0	0.00	0	0.00	3	0.67	3	0.26
77 - Refusal	13	4.18	55	13.51	82	18.34	150	12.88
78 - Parental Refusal	17	5.47	0	0.00	0	0.00	17	1.46
Other	2	0.64	19	4.67	3	0.67	24	2.06

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Oregon) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	311	100.00	407	100.00	447	100.00	1,165	100.00
70 - Interview Complete	270	87.03	318	79.10	335	75.05	923	76.48
71 - No One at DU	0	0.00	5	1.41	3	0.73	8	0.75
72 - Respondent Unavailable	5	1.87	8	1.72	14	2.83	27	2.62
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	1.20	2	0.49	5	1.68	11	1.51
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	0.43	2	0.35
76 - Language Barrier - Other	0	0.00	0	0.00	3	1.32	3	1.06
77 - Refusal	13	4.51	55	12.61	82	17.49	150	15.88
78 - Parental Refusal	17	4.91	0	0.00	0	0.00	17	0.39
Other	2	0.49	19	4.67	3	0.47	24	0.96

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Pennsylvania) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,425	100.00	1,536	100.00	1,744	100.00	4,705	100.00
70 - Interview Complete	1,169	82.04	1,218	79.30	1,193	68.41	3,580	76.09
71 - No One at DU	9	0.63	25	1.63	29	1.66	63	1.34
72 - Respondent Unavailable	13	0.91	34	2.21	14	0.80	61	1.30
73 - Break Off (Partial Interview)	0	0.00	1	0.07	0	0.00	1	0.02
74 - Physically/Mentally Incompetent	17	1.19	10	0.65	38	2.18	65	1.38
75 - Language Barrier - Spanish	0	0.00	1	0.07	5	0.29	6	0.13
76 - Language Barrier - Other	0	0.00	3	0.20	10	0.57	13	0.28
77 - Refusal	59	4.14	229	14.91	447	25.63	735	15.62
78 - Parental Refusal	149	10.46	0	0.00	0	0.00	149	3.17
Other	9	0.63	15	0.98	8	0.46	32	0.68

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Pennsylvania) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,425	100.00	1,536	100.00	1,744	100.00	4,705	100.00
70 - Interview Complete	1,169	82.15	1,218	79.74	1,193	67.91	3,580	70.67
71 - No One at DU	9	0.49	25	1.58	29	1.47	63	1.40
72 - Respondent Unavailable	13	0.98	34	2.25	14	0.60	61	0.84
73 - Break Off (Partial Interview)	0	0.00	1	0.03	0	0.00	1	0.00
74 - Physically/Mentally Incompetent	17	1.03	10	0.49	38	2.86	65	2.40
75 - Language Barrier - Spanish	0	0.00	1	0.06	5	0.33	6	0.27
76 - Language Barrier - Other	0	0.00	3	0.28	10	0.96	13	0.79
77 - Refusal	59	3.92	229	14.55	447	25.41	735	22.14
78 - Parental Refusal	149	10.82	0	0.00	0	0.00	149	0.94
Other	9	0.62	15	1.01	8	0.46	32	0.54

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Rhode Island) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	320	100.00	391	100.00	420	100.00	1,131	100.00
70 - Interview Complete	276	86.25	329	84.14	318	75.71	923	81.61
71 - No One at DU	1	0.31	4	1.02	2	0.48	7	0.62
72 - Respondent Unavailable	2	0.63	6	1.53	4	0.95	12	1.06
73 - Break Off (Partial Interview)	0	0.00	1	0.26	0	0.00	1	0.09
74 - Physically/Mentally Incompetent	1	0.31	3	0.77	6	1.43	10	0.88
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.31	0	0.00	3	0.71	4	0.35
77 - Refusal	9	2.81	43	11.00	85	20.24	137	12.11
78 - Parental Refusal	28	8.75	0	0.00	0	0.00	28	2.48
Other	2	0.63	5	1.28	2	0.48	9	0.80

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Rhode Island) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	320	100.00	391	100.00	420	100.00	1,131	100.00
70 - Interview Complete	276	86.40	329	84.47	318	75.55	923	77.76
71 - No One at DU	1	0.21	4	0.93	2	0.56	7	0.58
72 - Respondent Unavailable	2	0.62	6	1.42	4	0.99	12	1.02
73 - Break Off (Partial Interview)	0	0.00	1	0.29	0	0.00	1	0.04
74 - Physically/Mentally Incompetent	1	0.23	3	0.83	6	1.76	10	1.49
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.07	0	0.00	3	0.70	4	0.55
77 - Refusal	9	3.10	43	10.88	85	19.97	137	17.22
78 - Parental Refusal	28	9.11	0	0.00	0	0.00	28	0.77
Other	2	0.26	5	1.19	2	0.47	9	0.56

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (South Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	385	100.00	349	100.00	437	100.00	1,171	100.00
70 - Interview Complete	317	82.34	295	84.53	326	74.60	938	80.10
71 - No One at DU	2	0.52	10	2.87	5	1.14	17	1.45
72 - Respondent Unavailable	3	0.78	6	1.72	17	3.89	26	2.22
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.23	1	0.09
74 - Physically/Mentally Incompetent	4	1.04	5	1.43	9	2.06	18	1.54
75 - Language Barrier - Spanish	0	0.00	1	0.29	3	0.69	4	0.34
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.23	1	0.09
77 - Refusal	1	0.26	27	7.74	71	16.25	99	8.45
78 - Parental Refusal	53	13.77	0	0.00	0	0.00	53	4.53
Other	5	1.30	5	1.43	4	0.92	14	1.20

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (South Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	385	100.00	349	100.00	437	100.00	1,171	100.00
70 - Interview Complete	317	81.59	295	84.67	326	72.71	938	75.13
71 - No One at DU	2	0.52	10	2.64	5	1.15	17	1.29
72 - Respondent Unavailable	3	1.00	6	1.56	17	3.64	26	3.11
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.26	1	0.20
74 - Physically/Mentally Incompetent	4	1.03	5	1.68	9	2.22	18	2.03
75 - Language Barrier - Spanish	0	0.00	1	0.16	3	0.36	4	0.30
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.72	1	0.56
77 - Refusal	1	0.27	27	7.80	71	18.15	99	15.09
78 - Parental Refusal	53	14.52	0	0.00	0	0.00	53	1.37
Other	5	1.08	5	1.49	4	0.79	14	0.91

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (South Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	316	100.00	371	100.00	426	100.00	1,113	100.00
70 - Interview Complete	265	83.86	300	80.86	313	73.47	878	78.89
71 - No One at DU	5	1.58	8	2.16	10	2.35	23	2.07
72 - Respondent Unavailable	1	0.32	4	1.08	3	0.70	8	0.72
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.32	7	1.89	6	1.41	14	1.26
75 - Language Barrier - Spanish	0	0.00	3	0.81	4	0.94	7	0.63
76 - Language Barrier - Other	0	0.00	0	0.00	3	0.70	3	0.27
77 - Refusal	11	3.48	37	9.97	82	19.25	130	11.68
78 - Parental Refusal	29	9.18	0	0.00	0	0.00	29	2.61
Other	4	1.27	12	3.23	5	1.17	21	1.89

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (South Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	316	100.00	371	100.00	426	100.00	1,113	100.00
70 - Interview Complete	265	84.11	300	82.83	313	74.13	878	76.12
71 - No One at DU	5	2.18	8	1.46	10	1.69	23	1.70
72 - Respondent Unavailable	1	0.23	4	1.16	3	0.40	8	0.48
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	0.41	7	1.38	6	1.80	14	1.62
75 - Language Barrier - Spanish	0	0.00	3	0.66	4	0.84	7	0.74
76 - Language Barrier - Other	0	0.00	0	0.00	3	0.51	3	0.40
77 - Refusal	11	3.34	37	9.33	82	19.56	130	16.81
78 - Parental Refusal	29	8.50	0	0.00	0	0.00	29	0.76
Other	4	1.22	12	3.18	5	1.09	21	1.37

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Tennessee) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	299	100.00	419	100.00	387	100.00	1,105	100.00
70 - Interview Complete	261	87.29	352	84.01	314	81.14	927	83.89
71 - No One at DU	13	4.35	15	3.58	4	1.03	32	2.90
72 - Respondent Unavailable	0	0.00	5	1.19	5	1.29	10	0.90
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.67	2	0.48	2	0.52	6	0.54
75 - Language Barrier - Spanish	0	0.00	3	0.72	6	1.55	9	0.81
76 - Language Barrier - Other	0	0.00	0	0.00	2	0.52	2	0.18
77 - Refusal	3	1.00	36	8.59	54	13.95	93	8.42
78 - Parental Refusal	19	6.35	0	0.00	0	0.00	19	1.72
Other	1	0.33	6	1.43	0	0.00	7	0.63

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Tennessee) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	299	100.00	419	100.00	387	100.00	1,105	100.00
70 - Interview Complete	261	85.96	352	83.32	314	80.11	927	81.06
71 - No One at DU	13	4.50	15	3.68	4	0.91	32	1.60
72 - Respondent Unavailable	0	0.00	5	1.12	5	0.95	10	0.88
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.39	2	0.61	2	0.80	6	0.74
75 - Language Barrier - Spanish	0	0.00	3	0.91	6	1.18	9	1.04
76 - Language Barrier - Other	0	0.00	0	0.00	2	1.43	2	1.12
77 - Refusal	3	0.83	36	9.06	54	14.62	93	12.64
78 - Parental Refusal	19	7.96	0	0.00	0	0.00	19	0.72
Other	1	0.37	6	1.31	0	0.00	7	0.20

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Texas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,472	100.00	1,471	100.00	1,669	100.00	4,612	100.00
70 - Interview Complete	1,246	84.65	1,183	80.42	1,196	71.66	3,625	78.60
71 - No One at DU	19	1.29	52	3.54	68	4.07	139	3.01
72 - Respondent Unavailable	20	1.36	45	3.06	48	2.88	113	2.45
73 - Break Off (Partial Interview)	0	0.00	1	0.07	2	0.12	3	0.07
74 - Physically/Mentally Incompetent	15	1.02	12	0.82	21	1.26	48	1.04
75 - Language Barrier - Spanish	0	0.00	5	0.34	7	0.42	12	0.26
76 - Language Barrier - Other	0	0.00	1	0.07	11	0.66	12	0.26
77 - Refusal	30	2.04	145	9.86	305	18.27	480	10.41
78 - Parental Refusal	126	8.56	0	0.00	0	0.00	126	2.73
Other	16	1.09	27	1.84	11	0.66	54	1.17

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Texas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	1,472	100.00	1,471	100.00	1,669	100.00	4,612	100.00
70 - Interview Complete	1,246	84.37	1,183	80.38	1,196	70.45	3,625	73.36
71 - No One at DU	19	1.33	52	3.53	68	3.73	139	3.44
72 - Respondent Unavailable	20	1.31	45	2.92	48	2.94	113	2.76
73 - Break Off (Partial Interview)	0	0.00	1	0.05	2	0.14	3	0.11
74 - Physically/Mentally Incompetent	15	1.00	12	0.88	21	1.71	48	1.51
75 - Language Barrier - Spanish	0	0.00	5	0.23	7	0.47	12	0.38
76 - Language Barrier - Other	0	0.00	1	0.08	11	0.86	12	0.66
77 - Refusal	30	2.07	145	10.20	305	19.02	480	15.94
78 - Parental Refusal	126	8.86	0	0.00	0	0.00	126	0.97
Other	16	1.06	27	1.73	11	0.69	54	0.88

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Utah) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	319	100.00	384	100.00	396	100.00	1,099	100.00
70 - Interview Complete	287	89.97	310	80.73	329	83.08	926	84.26
71 - No One at DU	0	0.00	5	1.30	7	1.77	12	1.09
72 - Respondent Unavailable	6	1.88	12	3.13	10	2.53	28	2.55
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.25	1	0.09
74 - Physically/Mentally Incompetent	2	0.63	2	0.52	2	0.51	6	0.55
75 - Language Barrier - Spanish	0	0.00	3	0.78	0	0.00	3	0.27
76 - Language Barrier - Other	0	0.00	1	0.26	3	0.76	4	0.36
77 - Refusal	7	2.19	46	11.98	44	11.11	97	8.83
78 - Parental Refusal	16	5.02	0	0.00	0	0.00	16	1.46
Other	1	0.31	5	1.30	0	0.00	6	0.55

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Utah) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	319	100.00	384	100.00	396	100.00	1,099	100.00
70 - Interview Complete	287	90.49	310	81.78	329	82.34	926	83.26
71 - No One at DU	0	0.00	5	0.99	7	1.16	12	0.98
72 - Respondent Unavailable	6	2.19	12	3.11	10	2.62	28	2.65
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.19	1	0.13
74 - Physically/Mentally Incompetent	2	0.64	2	0.57	2	0.68	6	0.66
75 - Language Barrier - Spanish	0	0.00	3	0.70	0	0.00	3	0.12
76 - Language Barrier - Other	0	0.00	1	0.29	3	1.17	4	0.87
77 - Refusal	7	1.49	46	11.10	44	11.84	97	10.42
78 - Parental Refusal	16	4.90	0	0.00	0	0.00	16	0.61
Other	1	0.29	5	1.47	0	0.00	6	0.29

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Vermont) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	350	100.00	393	100.00	393	100.00	1,136	100.00
70 - Interview Complete	285	81.43	315	80.15	285	72.52	885	77.90
71 - No One at DU	4	1.14	7	1.78	5	1.27	16	1.41
72 - Respondent Unavailable	8	2.29	7	1.78	10	2.54	25	2.20
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	0.86	3	0.76	4	1.02	10	0.88
75 - Language Barrier - Spanish	1	0.29	0	0.00	0	0.00	1	0.09
76 - Language Barrier - Other	0	0.00	0	0.00	4	1.02	4	0.35
77 - Refusal	19	5.43	50	12.72	80	20.36	149	13.12
78 - Parental Refusal	28	8.00	0	0.00	0	0.00	28	2.46
Other	2	0.57	11	2.80	5	1.27	18	1.58

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Vermont) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	350	100.00	393	100.00	393	100.00	1,136	100.00
70 - Interview Complete	285	80.52	315	80.92	285	71.93	885	73.81
71 - No One at DU	4	0.98	7	1.48	5	1.26	16	1.26
72 - Respondent Unavailable	8	2.40	7	2.51	10	2.43	25	2.44
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	1.18	3	0.68	4	1.11	10	1.06
75 - Language Barrier - Spanish	1	0.19	0	0.00	0	0.00	1	0.02
76 - Language Barrier - Other	0	0.00	0	0.00	4	1.80	4	1.42
77 - Refusal	19	5.01	50	11.63	80	20.15	149	17.80
78 - Parental Refusal	28	9.44	0	0.00	0	0.00	28	0.76
Other	2	0.28	11	2.78	5	1.33	18	1.44

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	373	100.00	316	100.00	406	100.00	1,095	100.00
70 - Interview Complete	322	86.33	270	85.44	302	74.38	894	81.64
71 - No One at DU	0	0.00	1	0.32	4	0.99	5	0.46
72 - Respondent Unavailable	7	1.88	7	2.22	8	1.97	22	2.01
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.61	0	0.00	4	0.99	10	0.91
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.32	6	1.48	7	0.64
77 - Refusal	9	2.41	36	11.39	81	19.95	126	11.51
78 - Parental Refusal	27	7.24	0	0.00	0	0.00	27	2.47
Other	2	0.54	1	0.32	1	0.25	4	0.37

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	373	100.00	316	100.00	406	100.00	1,095	100.00
70 - Interview Complete	322	85.05	270	85.19	302	73.95	894	76.50
71 - No One at DU	0	0.00	1	0.33	4	0.78	5	0.65
72 - Respondent Unavailable	7	2.06	7	3.08	8	1.59	22	1.84
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	1.91	0	0.00	4	1.35	10	1.22
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.30	6	1.70	7	1.36
77 - Refusal	9	2.28	36	10.62	81	20.45	126	17.42
78 - Parental Refusal	27	7.92	0	0.00	0	0.00	27	0.74
Other	2	0.78	1	0.49	1	0.17	4	0.27

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Washington) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	368	100.00	406	100.00	444	100.00	1,218	100.00
70 - Interview Complete	301	81.79	310	76.35	317	71.40	928	76.19
71 - No One at DU	2	0.54	5	1.23	5	1.13	12	0.99
72 - Respondent Unavailable	5	1.36	14	3.45	10	2.25	29	2.38
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.54	2	0.49	6	1.35	10	0.82
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.23	1	0.08
76 - Language Barrier - Other	2	0.54	1	0.25	9	2.03	12	0.99
77 - Refusal	13	3.53	63	15.52	93	20.95	169	13.88
78 - Parental Refusal	42	11.41	0	0.00	0	0.00	42	3.45
Other	1	0.27	11	2.71	3	0.68	15	1.23

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Washington) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	368	100.00	406	100.00	444	100.00	1,218	100.00
70 - Interview Complete	301	81.58	310	75.50	317	70.24	928	71.82
71 - No One at DU	2	0.37	5	1.35	5	1.55	12	1.43
72 - Respondent Unavailable	5	1.60	14	3.48	10	2.91	29	2.87
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.50	2	1.10	6	1.98	10	1.75
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.29	1	0.23
76 - Language Barrier - Other	2	0.79	1	0.34	9	2.86	12	2.39
77 - Refusal	13	4.04	63	15.37	93	19.12	169	17.41
78 - Parental Refusal	42	10.89	0	0.00	0	0.00	42	0.91
Other	1	0.22	11	2.87	3	1.05	15	1.20

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (West Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	359	100.00	433	100.00	425	100.00	1,217	100.00
70 - Interview Complete	315	87.74	352	81.29	309	72.71	976	80.20
71 - No One at DU	7	1.95	17	3.93	11	2.59	35	2.88
72 - Respondent Unavailable	3	0.84	15	3.46	3	0.71	21	1.73
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	0.84	2	0.46	12	2.82	17	1.40
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	12	3.34	45	10.39	89	20.94	146	12.00
78 - Parental Refusal	17	4.74	0	0.00	0	0.00	17	1.40
Other	2	0.56	2	0.46	1	0.24	5	0.41

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (West Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	359	100.00	433	100.00	425	100.00	1,217	100.00
70 - Interview Complete	315	87.64	352	81.40	309	71.61	976	74.07
71 - No One at DU	7	1.77	17	3.60	11	2.14	35	2.28
72 - Respondent Unavailable	3	1.00	15	3.74	3	0.87	21	1.20
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	0.66	2	0.44	12	3.54	17	2.95
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	12	3.60	45	10.33	89	21.61	146	18.82
78 - Parental Refusal	17	4.67	0	0.00	0	0.00	17	0.39
Other	2	0.66	2	0.49	1	0.23	5	0.30

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Wisconsin) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	313	100.00	420	100.00	365	100.00	1,098	100.00
70 - Interview Complete	274	87.54	326	77.62	275	75.34	875	79.69
71 - No One at DU	1	0.32	6	1.43	4	1.10	11	1.00
72 - Respondent Unavailable	0	0.00	10	2.38	4	1.10	14	1.28
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.64	2	0.48	7	1.92	11	1.00
75 - Language Barrier - Spanish	0	0.00	2	0.48	4	1.10	6	0.55
76 - Language Barrier - Other	0	0.00	0	0.00	2	0.55	2	0.18
77 - Refusal	17	5.43	71	16.90	68	18.63	156	14.21
78 - Parental Refusal	18	5.75	0	0.00	0	0.00	18	1.64
Other	1	0.32	3	0.71	1	0.27	5	0.46

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Wisconsin) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	313	100.00	420	100.00	365	100.00	1,098	100.00
70 - Interview Complete	274	86.72	326	77.80	275	73.85	875	75.55
71 - No One at DU	1	0.28	6	1.26	4	0.93	11	0.91
72 - Respondent Unavailable	0	0.00	10	2.63	4	1.29	14	1.34
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	0.93	2	0.52	7	3.34	11	2.76
75 - Language Barrier - Spanish	0	0.00	2	0.20	4	0.84	6	0.68
76 - Language Barrier - Other	0	0.00	0	0.00	2	1.10	2	0.86
77 - Refusal	17	5.40	71	17.04	68	18.49	156	17.09
78 - Parental Refusal	18	6.43	0	0.00	0	0.00	18	0.60
Other	1	0.23	3	0.55	1	0.16	5	0.22

DU = dwelling unit.

Table 7.20 2012 Interview Results, by Age (Wyoming) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	337	100.00	358	100.00	453	100.00	1,148	100.00
70 - Interview Complete	293	86.94	285	79.61	350	77.26	928	80.84
71 - No One at DU	2	0.59	9	2.51	6	1.32	17	1.48
72 - Respondent Unavailable	4	1.19	15	4.19	10	2.21	29	2.53
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.48	2	0.56	8	1.77	15	1.31
75 - Language Barrier - Spanish	1	0.30	0	0.00	1	0.22	2	0.17
76 - Language Barrier - Other	0	0.00	1	0.28	2	0.44	3	0.26
77 - Refusal	8	2.37	41	11.45	71	15.67	120	10.45
78 - Parental Refusal	20	5.93	0	0.00	0	0.00	20	1.74
Other	4	1.19	5	1.40	5	1.10	14	1.22

DU = dwelling unit.

Table 7.21 2012 Interview Results, by Age (Wyoming) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Total Sample								
Eligible Cases	337	100.00	358	100.00	453	100.00	1,148	100.00
70 - Interview Complete	293	85.79	285	78.58	350	76.36	928	77.48
71 - No One at DU	2	0.65	9	2.77	6	1.02	17	1.20
72 - Respondent Unavailable	4	1.07	15	4.01	10	2.36	29	2.45
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	1.59	2	0.41	8	2.25	15	1.97
75 - Language Barrier - Spanish	1	0.31	0	0.00	1	0.16	2	0.15
76 - Language Barrier - Other	0	0.00	1	0.23	2	0.47	3	0.40
77 - Refusal	8	2.32	41	12.35	71	16.53	120	14.73
78 - Parental Refusal	20	7.14	0	0.00	0	0.00	20	0.65
Other	4	1.14	5	1.66	5	0.86	14	0.98

DU = dwelling unit.

Table 7.21a 2012 Interview Results, by Age and Race/Ethnicity (Total United States) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Hispanic or Latino								
Eligible Cases	4,936	100.00	4,954	100.00	4,016	100.00	13,906	100.00
70 - Interview Complete	4,294	86.97	3,945	79.67	2,929	71.38	11,168	74.95
71 - No One at DU	30	0.56	90	1.58	89	1.92	209	1.67
72 - Respondent Unavailable	52	0.91	138	2.86	134	3.44	324	3.00
73 - Break Off (Partial Interview)	0	0.00	3	0.07	3	0.10	6	0.08
74 - Physically/Mentally Incompetent	52	1.07	32	0.74	41	1.48	125	1.29
75 - Language Barrier - Spanish	7	0.09	38	0.60	64	1.33	109	1.04
76 - Language Barrier - Other	2	0.04	0	0.00	7	0.35	9	0.25
77 - Refusal	125	3.01	619	12.57	693	18.46	1,437	15.33
78 - Parental Refusal	326	6.45	0	0.00	0	0.00	326	0.87
Other	48	0.90	89	1.91	56	1.54	193	1.52
Not Hispanic or Latino								
Black or African American								
Eligible Cases	3,371	100.00	3,561	100.00	3,142	100.00	10,074	100.00
70 - Interview Complete	2,935	86.80	3,044	85.26	2,454	76.50	8,433	79.06
71 - No One at DU	24	0.66	50	1.43	52	1.57	126	1.44
72 - Respondent Unavailable	44	1.42	63	1.64	55	1.99	162	1.87
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.12	3	0.08
74 - Physically/Mentally Incompetent	30	1.26	22	0.74	67	2.78	119	2.29
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	4	0.15	6	0.12	21	0.60	31	0.47
77 - Refusal	86	2.53	310	8.96	461	15.60	857	13.06
78 - Parental Refusal	215	6.32	0	0.00	0	0.00	215	0.72
Other	33	0.85	66	1.85	29	0.84	128	1.00

Table 7.21a 2012 Interview Results, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino White								
Eligible Cases	16,436	100.00	17,650	100.00	22,288	100.00	56,374	100.00
70 - Interview Complete	13,316	80.69	13,829	78.12	16,020	70.47	43,165	72.19
71 - No One at DU	119	0.75	310	1.57	286	1.13	715	1.15
72 - Respondent Unavailable	152	0.90	348	1.92	338	1.43	838	1.44
73 - Break Off (Partial Interview)	0	0.00	5	0.03	6	0.05	11	0.04
74 - Physically/Mentally Incompetent	169	1.22	140	0.78	397	2.47	706	2.18
75 - Language Barrier - Spanish	2	0.00	1	0.00	1	0.00	4	0.00
76 - Language Barrier - Other	6	0.06	17	0.12	89	0.51	112	0.43
77 - Refusal	619	3.65	2,722	15.78	5,019	23.26	8,360	20.79
78 - Parental Refusal	1,966	12.07	0	0.00	0	0.00	1,966	0.99
Other	87	0.66	278	1.69	132	0.68	497	0.79
Not Hispanic or Latino American Indian or Alaska Native								
Eligible Cases	291	100.00	319	100.00	314	100.00	924	100.00
70 - Interview Complete	240	84.04	267	80.09	251	81.57	758	81.60
71 - No One at DU	3	0.63	5	0.92	3	0.18	11	0.34
72 - Respondent Unavailable	7	2.50	4	0.74	2	0.60	13	0.82
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	0.96	3	0.25	5	0.87	12	0.79
75 - Language Barrier - Spanish	0	0.00	1	0.21	0	0.00	1	0.03
76 - Language Barrier - Other	0	0.00	0	0.00	3	1.78	3	1.34
77 - Refusal	8	2.22	35	16.36	46	12.31	89	11.87
78 - Parental Refusal	27	9.42	0	0.00	0	0.00	27	0.96
Other	2	0.23	4	1.43	4	2.68	10	2.25

Table 7.21a 2012 Interview Results, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino Native Hawaiian or Other Pacific Islander								
Eligible Cases	111	100.00	163	100.00	154	100.00	428	100.00
70 - Interview Complete	92	91.07	138	80.92	125	79.65	355	80.84
71 - No One at DU	4	0.89	1	0.51	1	0.14	6	0.26
72 - Respondent Unavailable	3	0.53	3	1.06	3	2.60	9	2.16
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	0	0.00	0	0.00	1	0.15	1	0.12
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.23	1	1.90	1	0.55	3	0.75
77 - Refusal	2	0.34	13	9.73	22	16.70	37	14.15
78 - Parental Refusal	9	6.95	0	0.00	0	0.00	9	0.59
Other	0	0.00	7	5.88	1	0.20	8	1.13
Not Hispanic or Latino Asian								
Eligible Cases	887	100.00	1,155	100.00	1,428	100.00	3,470	100.00
70 - Interview Complete	684	75.80	848	71.41	858	56.50	2,390	60.04
71 - No One at DU	7	0.35	22	1.39	23	1.34	52	1.26
72 - Respondent Unavailable	7	1.11	27	2.39	28	1.65	62	1.70
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	0.76	9	1.02	13	0.98	29	0.97
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	0.07	2	0.06
76 - Language Barrier - Other	16	1.82	46	4.20	195	19.88	257	16.35
77 - Refusal	32	5.01	178	17.33	297	18.93	507	17.54
78 - Parental Refusal	125	14.46	0	0.00	0	0.00	125	1.23
Other	9	0.70	25	2.27	12	0.65	46	0.86

Table 7.21a 2012 Interview Results, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino Multiple Races								
Eligible Cases	1,115	100.00	837	100.00	528	100.00	2,480	100.00
70 - Interview Complete	931	82.88	691	83.46	418	80.49	2,040	81.62
71 - No One at DU	5	0.60	14	1.69	9	1.36	28	1.27
72 - Respondent Unavailable	11	1.07	15	1.01	11	2.29	37	1.76
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.12	1	0.07
74 - Physically/Mentally Incompetent	12	0.96	10	1.00	9	0.72	31	0.83
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.01	1	0.12	2	0.82	4	0.50
77 - Refusal	28	3.03	99	11.63	74	13.73	201	11.06
78 - Parental Refusal	119	10.76	0	0.00	0	0.00	119	2.24
Other	8	0.68	7	1.09	4	0.49	19	0.65

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Total United States) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	4,655	100.00	5,877	100.00	8,815	100.00	19,347	100.00
71 - No One at DU	192	4.12	492	8.37	463	5.25	1,147	5.93
72 - Respondent Unavailable	276	5.93	598	10.18	571	6.48	1,445	7.47
73 - Break Off (Partial Interview)	0	0.00	8	0.14	13	0.15	21	0.11
74 - Physically/Mentally Incompetent	274	5.89	216	3.68	533	6.05	1,023	5.29
75 - Language Barrier - Spanish	9	0.19	40	0.68	67	0.76	116	0.60
76 - Language Barrier - Other	30	0.64	71	1.21	318	3.61	419	2.17
77 - Refusal	900	19.33	3,976	67.65	6,612	75.01	11,488	59.38
78 - Parental Refusal	2,787	59.87	0	0.00	0	0.00	2,787	14.41
Other	187	4.02	476	8.10	238	2.70	901	4.66

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Total United States) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	4,655	100.00	5,877	100.00	8,815	100.00	19,347	100.00
71 - No One at DU	192	3.91	492	7.39	463	4.41	1,147	4.69
72 - Respondent Unavailable	276	5.85	598	9.96	571	6.12	1,445	6.50
73 - Break Off (Partial Interview)	0	0.00	8	0.14	13	0.20	21	0.18
74 - Physically/Mentally Incompetent	274	6.75	216	3.72	533	7.70	1,023	7.24
75 - Language Barrier - Spanish	9	0.12	40	0.62	67	0.65	116	0.61
76 - Language Barrier - Other	30	0.86	71	1.47	318	5.26	419	4.61
77 - Refusal	900	19.65	3,976	68.05	6,612	72.82	11,488	69.08
78 - Parental Refusal	2,787	58.60	0	0.00	0	0.00	2,787	3.58
Other	187	4.26	476	8.64	238	2.83	901	3.51

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Alabama) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	64	100.00	71	100.00	109	100.00	244	100.00
71 - No One at DU	2	3.13	5	7.04	3	2.75	10	4.10
72 - Respondent Unavailable	8	12.50	12	16.90	13	11.93	33	13.52
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	6.25	2	2.82	14	12.84	20	8.20
75 - Language Barrier - Spanish	2	3.13	2	2.82	4	3.67	8	3.28
76 - Language Barrier - Other	0	0.00	0	0.00	2	1.83	2	0.82
77 - Refusal	16	25.00	40	56.34	73	66.97	129	52.87
78 - Parental Refusal	31	48.44	0	0.00	0	0.00	31	12.70
Other	1	1.56	10	14.08	0	0.00	11	4.51

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Alabama) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	64	100.00	71	100.00	109	100.00	244	100.00
71 - No One at DU	2	2.61	5	6.01	3	1.57	10	2.11
72 - Respondent Unavailable	8	9.79	12	14.71	13	10.31	33	10.72
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	7.58	2	2.92	14	12.82	20	11.39
75 - Language Barrier - Spanish	2	2.30	2	3.36	4	2.08	8	2.23
76 - Language Barrier - Other	0	0.00	0	0.00	2	2.12	2	1.74
77 - Refusal	16	25.25	40	60.21	73	71.09	129	66.37
78 - Parental Refusal	31	51.43	0	0.00	0	0.00	31	4.05
Other	1	1.03	10	12.80	0	0.00	11	1.39

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Alaska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	62	100.00	114	100.00	247	100.00
71 - No One at DU	5	7.04	10	16.13	8	7.02	23	9.31
72 - Respondent Unavailable	12	16.90	9	14.52	6	5.26	27	10.93
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.88	1	0.40
74 - Physically/Mentally Incompetent	4	5.63	3	4.84	6	5.26	13	5.26
75 - Language Barrier - Spanish	2	2.82	0	0.00	2	1.75	4	1.62
76 - Language Barrier - Other	0	0.00	1	1.61	4	3.51	5	2.02
77 - Refusal	17	23.94	35	56.45	83	72.81	135	54.66
78 - Parental Refusal	30	42.25	0	0.00	0	0.00	30	12.15
Other	1	1.41	4	6.45	4	3.51	9	3.64

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Alaska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	62	100.00	114	100.00	247	100.00
71 - No One at DU	5	7.68	10	13.90	8	6.71	23	7.43
72 - Respondent Unavailable	12	19.38	9	14.79	6	5.66	27	7.66
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	1.71	1	1.41
74 - Physically/Mentally Incompetent	4	5.68	3	6.00	6	9.57	13	8.92
75 - Language Barrier - Spanish	2	1.84	0	0.00	2	1.70	4	1.56
76 - Language Barrier - Other	0	0.00	1	0.87	4	3.34	5	2.83
77 - Refusal	17	20.08	35	58.75	83	68.21	135	63.24
78 - Parental Refusal	30	43.80	0	0.00	0	0.00	30	3.75
Other	1	1.53	4	5.71	4	3.10	9	3.20

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Arizona) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	54	100.00	78	100.00	85	100.00	217	100.00
71 - No One at DU	2	3.70	9	11.54	5	5.88	16	7.37
72 - Respondent Unavailable	5	9.26	9	11.54	10	11.76	24	11.06
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	3.70	1	1.28	7	8.24	10	4.61
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	5	9.26	6	7.69	6	7.06	17	7.83
77 - Refusal	20	37.04	47	60.26	55	64.71	122	56.22
78 - Parental Refusal	15	27.78	0	0.00	0	0.00	15	6.91
Other	5	9.26	6	7.69	2	2.35	13	5.99

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Arizona) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	54	100.00	78	100.00	85	100.00	217	100.00
71 - No One at DU	2	3.37	9	8.07	5	7.19	16	7.08
72 - Respondent Unavailable	5	9.27	9	11.34	10	9.04	24	9.39
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	4.02	1	2.57	7	8.40	10	7.27
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	5	12.15	6	9.10	6	8.91	17	9.14
77 - Refusal	20	38.12	47	62.84	55	62.99	122	61.43
78 - Parental Refusal	15	25.69	0	0.00	0	0.00	15	1.59
Other	5	7.38	6	6.08	2	3.48	13	4.10

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Arkansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	82	100.00	94	100.00	123	100.00	299	100.00
71 - No One at DU	2	2.44	1	1.06	3	2.44	6	2.01
72 - Respondent Unavailable	0	0.00	4	4.26	4	3.25	8	2.68
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	7.32	1	1.06	6	4.88	13	4.35
75 - Language Barrier - Spanish	1	1.22	1	1.06	1	0.81	3	1.00
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.81	1	0.33
77 - Refusal	22	26.83	79	84.04	105	85.37	206	68.90
78 - Parental Refusal	48	58.54	0	0.00	0	0.00	48	16.05
Other	3	3.66	8	8.51	3	2.44	14	4.68

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Arkansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	82	100.00	94	100.00	123	100.00	299	100.00
71 - No One at DU	2	2.02	1	0.90	3	2.48	6	2.27
72 - Respondent Unavailable	0	0.00	4	3.75	4	2.48	8	2.45
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	7.34	1	0.78	6	6.37	13	5.81
75 - Language Barrier - Spanish	1	0.81	1	0.91	1	0.41	3	0.49
76 - Language Barrier - Other	0	0.00	0	0.00	1	3.32	1	2.72
77 - Refusal	22	29.49	79	85.16	105	82.84	206	79.42
78 - Parental Refusal	48	57.47	0	0.00	0	0.00	48	3.97
Other	3	2.86	8	8.49	3	2.10	14	2.87

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (California) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	250	100.00	368	100.00	553	100.00	1,171	100.00
71 - No One at DU	3	1.20	6	1.63	11	1.99	20	1.71
72 - Respondent Unavailable	8	3.20	14	3.80	17	3.07	39	3.33
73 - Break Off (Partial Interview)	0	0.00	1	0.27	1	0.18	2	0.17
74 - Physically/Mentally Incompetent	11	4.40	14	3.80	31	5.61	56	4.78
75 - Language Barrier - Spanish	1	0.40	0	0.00	0	0.00	1	0.09
76 - Language Barrier - Other	3	1.20	11	2.99	44	7.96	58	4.95
77 - Refusal	70	28.00	296	80.43	425	76.85	791	67.55
78 - Parental Refusal	145	58.00	0	0.00	0	0.00	145	12.38
Other	9	3.60	26	7.07	24	4.34	59	5.04

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (California) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	250	100.00	368	100.00	553	100.00	1,171	100.00
71 - No One at DU	3	1.24	6	1.52	11	1.76	20	1.70
72 - Respondent Unavailable	8	3.29	14	4.39	17	2.43	39	2.70
73 - Break Off (Partial Interview)	0	0.00	1	0.20	1	0.23	2	0.22
74 - Physically/Mentally Incompetent	11	5.13	14	3.35	31	6.70	56	6.24
75 - Language Barrier - Spanish	1	0.38	0	0.00	0	0.00	1	0.02
76 - Language Barrier - Other	3	1.11	11	2.42	44	9.52	58	8.23
77 - Refusal	70	31.21	296	80.32	425	75.45	791	73.28
78 - Parental Refusal	145	54.09	0	0.00	0	0.00	145	3.31
Other	9	3.56	26	7.80	24	3.90	59	4.31

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Colorado) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	57	100.00	89	100.00	115	100.00	261	100.00
71 - No One at DU	0	0.00	2	2.25	0	0.00	2	0.77
72 - Respondent Unavailable	6	10.53	14	15.73	22	19.13	42	16.09
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.75	5	5.62	2	1.74	8	3.07
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	1.12	4	3.48	5	1.92
77 - Refusal	10	17.54	58	65.17	86	74.78	154	59.00
78 - Parental Refusal	36	63.16	0	0.00	0	0.00	36	13.79
Other	4	7.02	9	10.11	1	0.87	14	5.36

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Colorado) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	57	100.00	89	100.00	115	100.00	261	100.00
71 - No One at DU	0	0.00	2	1.66	0	0.00	2	0.19
72 - Respondent Unavailable	6	11.72	14	15.69	22	17.56	42	17.06
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.63	5	3.94	2	2.37	8	2.52
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	0.69	4	4.28	5	3.66
77 - Refusal	10	19.69	58	67.71	86	75.36	154	71.71
78 - Parental Refusal	36	59.91	0	0.00	0	0.00	36	2.97
Other	4	7.04	9	10.31	1	0.42	14	1.89

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Connecticut) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	73	100.00	87	100.00	137	100.00	297	100.00
71 - No One at DU	1	1.37	7	8.05	8	5.84	16	5.39
72 - Respondent Unavailable	1	1.37	7	8.05	12	8.76	20	6.73
73 - Break Off (Partial Interview)	0	0.00	1	1.15	0	0.00	1	0.34
74 - Physically/Mentally Incompetent	1	1.37	1	1.15	13	9.49	15	5.05
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	1.37	0	0.00	3	2.19	4	1.35
77 - Refusal	13	17.81	65	74.71	98	71.53	176	59.26
78 - Parental Refusal	51	69.86	0	0.00	0	0.00	51	17.17
Other	5	6.85	6	6.90	3	2.19	14	4.71

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Connecticut) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	73	100.00	87	100.00	137	100.00	297	100.00
71 - No One at DU	1	0.96	7	7.97	8	3.30	16	3.53
72 - Respondent Unavailable	1	1.15	7	9.76	12	6.83	20	6.71
73 - Break Off (Partial Interview)	0	0.00	1	1.03	0	0.00	1	0.08
74 - Physically/Mentally Incompetent	1	1.66	1	0.99	13	11.74	15	10.24
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	1.14	0	0.00	3	4.15	4	3.62
77 - Refusal	13	18.93	65	71.80	98	72.12	176	68.76
78 - Parental Refusal	51	70.35	0	0.00	0	0.00	51	4.41
Other	5	5.81	6	8.45	3	1.87	14	2.65

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Delaware) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	69	100.00	59	100.00	89	100.00	217	100.00
71 - No One at DU	0	0.00	1	1.69	2	2.25	3	1.38
72 - Respondent Unavailable	0	0.00	2	3.39	2	2.25	4	1.84
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	2.90	4	6.78	4	4.49	10	4.61
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	1	1.12	1	0.46
77 - Refusal	10	14.49	48	81.36	80	89.89	138	63.59
78 - Parental Refusal	55	79.71	0	0.00	0	0.00	55	25.35
Other	2	2.90	4	6.78	0	0.00	6	2.76

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Delaware) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	69	100.00	59	100.00	89	100.00	217	100.00
71 - No One at DU	0	0.00	1	1.22	2	2.38	3	2.10
72 - Respondent Unavailable	0	0.00	2	4.02	2	2.30	4	2.31
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	2.88	4	4.09	4	4.71	10	4.52
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.74	1	0.61
77 - Refusal	10	13.70	48	81.03	80	89.86	138	83.61
78 - Parental Refusal	55	80.24	0	0.00	0	0.00	55	5.69
Other	2	3.17	4	9.64	0	0.00	6	1.16

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (District of Columbia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	33	100.00	54	100.00	76	100.00	163	100.00
71 - No One at DU	0	0.00	6	11.11	2	2.63	8	4.91
72 - Respondent Unavailable	3	9.09	7	12.96	7	9.21	17	10.43
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	6.06	0	0.00	8	10.53	10	6.13
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	3.03	0	0.00	4	5.26	5	3.07
77 - Refusal	7	21.21	40	74.07	53	69.74	100	61.35
78 - Parental Refusal	18	54.55	0	0.00	0	0.00	18	11.04
Other	2	6.06	1	1.85	2	2.63	5	3.07

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (District of Columbia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	33	100.00	54	100.00	76	100.00	163	100.00
71 - No One at DU	0	0.00	6	10.23	2	3.19	8	3.94
72 - Respondent Unavailable	3	8.82	7	16.11	7	9.38	17	10.15
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	5.83	0	0.00	8	11.09	10	9.68
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	3.69	0	0.00	4	4.69	5	4.12
77 - Refusal	7	23.22	40	71.29	53	68.56	100	67.80
78 - Parental Refusal	18	53.40	0	0.00	0	0.00	18	1.27
Other	2	5.03	1	2.36	2	3.08	5	3.04

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Florida) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	226	100.00	313	100.00	496	100.00	1,035	100.00
71 - No One at DU	9	3.98	9	2.88	7	1.41	25	2.42
72 - Respondent Unavailable	16	7.08	24	7.67	38	7.66	78	7.54
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.60	3	0.29
74 - Physically/Mentally Incompetent	17	7.52	14	4.47	33	6.65	64	6.18
75 - Language Barrier - Spanish	0	0.00	1	0.32	0	0.00	1	0.10
76 - Language Barrier - Other	2	0.88	3	0.96	17	3.43	22	2.13
77 - Refusal	29	12.83	217	69.33	359	72.38	605	58.45
78 - Parental Refusal	131	57.96	0	0.00	0	0.00	131	12.66
Other	22	9.73	45	14.38	39	7.86	106	10.24

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Florida) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	226	100.00	313	100.00	496	100.00	1,035	100.00
71 - No One at DU	9	3.84	9	2.43	7	1.32	25	1.54
72 - Respondent Unavailable	16	8.21	24	7.23	38	6.46	78	6.61
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.58	3	0.50
74 - Physically/Mentally Incompetent	17	7.27	14	6.88	33	8.58	64	8.37
75 - Language Barrier - Spanish	0	0.00	1	0.19	0	0.00	1	0.02
76 - Language Barrier - Other	2	0.78	3	0.81	17	4.02	22	3.58
77 - Refusal	29	11.57	217	68.60	359	72.65	605	69.31
78 - Parental Refusal	131	58.57	0	0.00	0	0.00	131	2.86
Other	22	9.76	45	13.87	39	6.39	106	7.20

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Georgia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	57	100.00	76	100.00	126	100.00	259	100.00
71 - No One at DU	0	0.00	0	0.00	2	1.59	2	0.77
72 - Respondent Unavailable	1	1.75	9	11.84	8	6.35	18	6.95
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.79	1	0.39
74 - Physically/Mentally Incompetent	9	15.79	2	2.63	9	7.14	20	7.72
75 - Language Barrier - Spanish	0	0.00	2	2.63	5	3.97	7	2.70
76 - Language Barrier - Other	0	0.00	0	0.00	6	4.76	6	2.32
77 - Refusal	22	38.60	53	69.74	86	68.25	161	62.16
78 - Parental Refusal	17	29.82	0	0.00	0	0.00	17	6.56
Other	8	14.04	10	13.16	9	7.14	27	10.42

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Georgia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	57	100.00	76	100.00	126	100.00	259	100.00
71 - No One at DU	0	0.00	0	0.00	2	2.31	2	1.92
72 - Respondent Unavailable	1	0.91	9	11.95	8	5.70	18	6.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.54	1	0.45
74 - Physically/Mentally Incompetent	9	18.97	2	3.21	9	7.09	20	7.51
75 - Language Barrier - Spanish	0	0.00	2	2.69	5	2.43	7	2.29
76 - Language Barrier - Other	0	0.00	0	0.00	6	5.32	6	4.42
77 - Refusal	22	31.84	53	66.40	86	67.44	161	64.91
78 - Parental Refusal	17	35.49	0	0.00	0	0.00	17	2.42
Other	8	12.79	10	15.74	9	9.16	27	10.07

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Hawaii) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	93	100.00	74	100.00	180	100.00	347	100.00
71 - No One at DU	8	8.60	9	12.16	14	7.78	31	8.93
72 - Respondent Unavailable	3	3.23	5	6.76	7	3.89	15	4.32
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	5.38	2	2.70	11	6.11	18	5.19
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.56	1	0.29
76 - Language Barrier - Other	4	4.30	1	1.35	18	10.00	23	6.63
77 - Refusal	11	11.83	52	70.27	124	68.89	187	53.89
78 - Parental Refusal	61	65.59	0	0.00	0	0.00	61	17.58
Other	1	1.08	5	6.76	5	2.78	11	3.17

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Hawaii) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	93	100.00	74	100.00	180	100.00	347	100.00
71 - No One at DU	8	6.25	9	12.88	14	6.35	31	6.82
72 - Respondent Unavailable	3	1.99	5	6.76	7	4.16	15	4.22
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	6.49	2	2.08	11	5.74	18	5.52
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.89	1	0.77
76 - Language Barrier - Other	4	3.77	1	0.91	18	11.39	23	10.14
77 - Refusal	11	13.59	52	70.95	124	68.76	187	65.44
78 - Parental Refusal	61	66.94	0	0.00	0	0.00	61	4.22
Other	1	0.97	5	6.42	5	2.70	11	2.87

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Idaho) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	72	100.00	99	100.00	215	100.00
71 - No One at DU	0	0.00	1	1.39	3	3.03	4	1.86
72 - Respondent Unavailable	2	4.55	3	4.17	1	1.01	6	2.79
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	9.09	1	1.39	1	1.01	6	2.79
75 - Language Barrier - Spanish	0	0.00	2	2.78	4	4.04	6	2.79
76 - Language Barrier - Other	0	0.00	0	0.00	1	1.01	1	0.47
77 - Refusal	15	34.09	63	87.50	89	89.90	167	77.67
78 - Parental Refusal	20	45.45	0	0.00	0	0.00	20	9.30
Other	3	6.82	2	2.78	0	0.00	5	2.33

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Idaho) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	72	100.00	99	100.00	215	100.00
71 - No One at DU	0	0.00	1	1.20	3	2.70	4	2.32
72 - Respondent Unavailable	2	3.76	3	3.21	1	1.02	6	1.49
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	7.57	1	1.09	1	1.91	6	2.15
75 - Language Barrier - Spanish	0	0.00	2	2.11	4	3.78	6	3.31
76 - Language Barrier - Other	0	0.00	0	0.00	1	2.39	1	1.92
77 - Refusal	15	35.75	63	89.98	89	88.20	167	85.15
78 - Parental Refusal	20	47.22	0	0.00	0	0.00	20	2.97
Other	3	5.69	2	2.41	0	0.00	5	0.69

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Illinois) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	283	100.00	372	100.00	544	100.00	1,199	100.00
71 - No One at DU	18	6.36	60	16.13	57	10.48	135	11.26
72 - Respondent Unavailable	15	5.30	30	8.06	32	5.88	77	6.42
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	15	5.30	19	5.11	32	5.88	66	5.50
75 - Language Barrier - Spanish	0	0.00	1	0.27	0	0.00	1	0.08
76 - Language Barrier - Other	2	0.71	3	0.81	35	6.43	40	3.34
77 - Refusal	71	25.09	240	64.52	379	69.67	690	57.55
78 - Parental Refusal	154	54.42	0	0.00	0	0.00	154	12.84
Other	8	2.83	19	5.11	9	1.65	36	3.00

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Illinois) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	283	100.00	372	100.00	544	100.00	1,199	100.00
71 - No One at DU	18	5.92	60	15.98	57	8.86	135	9.42
72 - Respondent Unavailable	15	5.64	30	8.00	32	6.12	77	6.29
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	15	4.66	19	5.42	32	8.01	66	7.53
75 - Language Barrier - Spanish	0	0.00	1	0.20	0	0.00	1	0.02
76 - Language Barrier - Other	2	1.49	3	1.40	35	6.85	40	5.95
77 - Refusal	71	24.89	240	62.35	379	67.88	690	64.66
78 - Parental Refusal	154	54.05	0	0.00	0	0.00	154	3.33
Other	8	3.34	19	6.66	9	2.28	36	2.81

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Indiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	59	100.00	80	100.00	121	100.00	260	100.00
71 - No One at DU	4	6.78	6	7.50	4	3.31	14	5.38
72 - Respondent Unavailable	1	1.69	7	8.75	8	6.61	16	6.15
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	8.47	6	7.50	8	6.61	19	7.31
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	1.25	0	0.00	1	0.38
77 - Refusal	5	8.47	60	75.00	99	81.82	164	63.08
78 - Parental Refusal	43	72.88	0	0.00	0	0.00	43	16.54
Other	1	1.69	0	0.00	2	1.65	3	1.15

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Indiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	59	100.00	80	100.00	121	100.00	260	100.00
71 - No One at DU	4	6.68	6	7.46	4	3.32	14	3.89
72 - Respondent Unavailable	1	1.68	7	9.73	8	7.08	16	6.99
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	12.05	6	6.43	8	6.93	19	7.19
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	2.38	0	0.00	1	0.21
77 - Refusal	5	8.49	60	74.01	99	80.72	164	75.75
78 - Parental Refusal	43	69.71	0	0.00	0	0.00	43	4.21
Other	1	1.40	0	0.00	2	1.96	3	1.75

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Iowa) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	59	100.00	75	100.00	103	100.00	237	100.00
71 - No One at DU	1	1.69	8	10.67	0	0.00	9	3.80
72 - Respondent Unavailable	2	3.39	9	12.00	6	5.83	17	7.17
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	8.47	4	5.33	4	3.88	13	5.49
75 - Language Barrier - Spanish	0	0.00	1	1.33	0	0.00	1	0.42
76 - Language Barrier - Other	0	0.00	2	2.67	1	0.97	3	1.27
77 - Refusal	8	13.56	42	56.00	85	82.52	135	56.96
78 - Parental Refusal	42	71.19	0	0.00	0	0.00	42	17.72
Other	1	1.69	9	12.00	7	6.80	17	7.17

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Iowa) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	59	100.00	75	100.00	103	100.00	237	100.00
71 - No One at DU	1	0.93	8	10.71	0	0.00	9	1.28
72 - Respondent Unavailable	2	2.77	9	11.99	6	7.37	17	7.56
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	21.79	4	5.84	4	4.01	13	5.50
75 - Language Barrier - Spanish	0	0.00	1	0.96	0	0.00	1	0.11
76 - Language Barrier - Other	0	0.00	2	1.10	1	0.56	3	0.58
77 - Refusal	8	11.41	42	57.86	85	82.90	135	74.92
78 - Parental Refusal	42	61.91	0	0.00	0	0.00	42	4.47
Other	1	1.19	9	11.54	7	5.16	17	5.59

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Kansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	45	100.00	53	100.00	99	100.00	197	100.00
71 - No One at DU	1	2.22	0	0.00	6	6.06	7	3.55
72 - Respondent Unavailable	3	6.67	5	9.43	6	6.06	14	7.11
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	4.44	4	7.55	4	4.04	10	5.08
75 - Language Barrier - Spanish	0	0.00	1	1.89	1	1.01	2	1.02
76 - Language Barrier - Other	0	0.00	0	0.00	4	4.04	4	2.03
77 - Refusal	13	28.89	34	64.15	77	77.78	124	62.94
78 - Parental Refusal	24	53.33	0	0.00	0	0.00	24	12.18
Other	2	4.44	9	16.98	1	1.01	12	6.09

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Kansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	45	100.00	53	100.00	99	100.00	197	100.00
71 - No One at DU	1	2.33	0	0.00	6	5.79	7	5.05
72 - Respondent Unavailable	3	6.09	5	9.99	6	5.58	14	6.02
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	4.80	4	7.44	4	4.70	10	4.97
75 - Language Barrier - Spanish	0	0.00	1	1.66	1	1.57	2	1.49
76 - Language Barrier - Other	0	0.00	0	0.00	4	3.81	4	3.24
77 - Refusal	13	26.73	34	61.12	77	77.99	124	73.52
78 - Parental Refusal	24	55.97	0	0.00	0	0.00	24	3.16
Other	2	4.07	9	19.80	1	0.57	12	2.57

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Kentucky) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	66	100.00	78	100.00	113	100.00	257	100.00
71 - No One at DU	3	4.55	8	10.26	6	5.31	17	6.61
72 - Respondent Unavailable	6	9.09	10	12.82	6	5.31	22	8.56
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	7.58	4	5.13	12	10.62	21	8.17
75 - Language Barrier - Spanish	0	0.00	2	2.56	2	1.77	4	1.56
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.88	1	0.39
77 - Refusal	13	19.70	43	55.13	82	72.57	138	53.70
78 - Parental Refusal	36	54.55	0	0.00	0	0.00	36	14.01
Other	3	4.55	11	14.10	4	3.54	18	7.00

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Kentucky) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	66	100.00	78	100.00	113	100.00	257	100.00
71 - No One at DU	3	3.61	8	9.41	6	4.43	17	4.84
72 - Respondent Unavailable	6	8.81	10	11.20	6	3.57	22	4.62
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	6.73	4	4.08	12	12.52	21	11.36
75 - Language Barrier - Spanish	0	0.00	2	1.60	2	0.68	4	0.72
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.68	1	0.57
77 - Refusal	13	20.40	43	51.15	82	74.79	138	69.07
78 - Parental Refusal	36	55.43	0	0.00	0	0.00	36	3.57
Other	3	5.02	11	22.56	4	3.33	18	5.24

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Louisiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	38	100.00	61	100.00	100	100.00	199	100.00
71 - No One at DU	2	5.26	11	18.03	5	5.00	18	9.05
72 - Respondent Unavailable	3	7.89	4	6.56	7	7.00	14	7.04
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	7.89	5	8.20	8	8.00	16	8.04
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	1.64	1	1.00	2	1.01
77 - Refusal	5	13.16	36	59.02	76	76.00	117	58.79
78 - Parental Refusal	24	63.16	0	0.00	0	0.00	24	12.06
Other	1	2.63	4	6.56	3	3.00	8	4.02

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Louisiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	38	100.00	61	100.00	100	100.00	199	100.00
71 - No One at DU	2	8.94	11	21.39	5	5.35	18	7.30
72 - Respondent Unavailable	3	4.68	4	5.86	7	4.71	14	4.84
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	7.35	5	7.41	8	10.13	16	9.69
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	2.24	1	2.42	2	2.28
77 - Refusal	5	13.62	36	55.16	76	74.46	117	69.32
78 - Parental Refusal	24	63.48	0	0.00	0	0.00	24	3.14
Other	1	1.93	4	7.94	3	2.93	8	3.44

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Maine) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	54	100.00	62	100.00	80	100.00	196	100.00
71 - No One at DU	4	7.41	5	8.06	4	5.00	13	6.63
72 - Respondent Unavailable	4	7.41	3	4.84	3	3.75	10	5.10
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	3.70	5	8.06	8	10.00	15	7.65
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	2	3.23	1	1.25	3	1.53
77 - Refusal	10	18.52	44	70.97	63	78.75	117	59.69
78 - Parental Refusal	34	62.96	0	0.00	0	0.00	34	17.35
Other	0	0.00	3	4.84	1	1.25	4	2.04

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Maine) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	54	100.00	62	100.00	80	100.00	196	100.00
71 - No One at DU	4	8.17	5	8.01	4	3.17	13	3.92
72 - Respondent Unavailable	4	6.53	3	5.21	3	4.65	10	4.82
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	5.76	5	9.93	8	10.65	15	10.27
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	2	3.87	1	4.15	3	3.86
77 - Refusal	10	18.58	44	65.47	63	76.55	117	71.85
78 - Parental Refusal	34	60.96	0	0.00	0	0.00	34	3.92
Other	0	0.00	3	7.52	1	0.83	4	1.37

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Maryland) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	48	100.00	57	100.00	95	100.00	200	100.00
71 - No One at DU	0	0.00	2	3.51	0	0.00	2	1.00
72 - Respondent Unavailable	3	6.25	6	10.53	1	1.05	10	5.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	2.08	2	3.51	11	11.58	14	7.00
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	1.75	6	6.32	7	3.50
77 - Refusal	8	16.67	44	77.19	76	80.00	128	64.00
78 - Parental Refusal	35	72.92	0	0.00	0	0.00	35	17.50
Other	1	2.08	2	3.51	1	1.05	4	2.00

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Maryland) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	48	100.00	57	100.00	95	100.00	200	100.00
71 - No One at DU	0	0.00	2	2.51	0	0.00	2	0.23
72 - Respondent Unavailable	3	4.90	6	10.63	1	0.72	10	1.88
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.63	2	2.18	11	14.17	14	12.32
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	1.29	6	6.04	7	5.24
77 - Refusal	8	14.68	44	80.42	76	78.39	128	74.80
78 - Parental Refusal	35	76.38	0	0.00	0	0.00	35	4.53
Other	1	2.41	2	2.97	1	0.69	4	1.00

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Massachusetts) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	96	100.00	131	100.00	298	100.00
71 - No One at DU	0	0.00	7	7.29	6	4.58	13	4.36
72 - Respondent Unavailable	5	7.04	21	21.88	11	8.40	37	12.42
73 - Break Off (Partial Interview)	0	0.00	1	1.04	1	0.76	2	0.67
74 - Physically/Mentally Incompetent	4	5.63	3	3.13	9	6.87	16	5.37
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.76	1	0.34
76 - Language Barrier - Other	0	0.00	4	4.17	13	9.92	17	5.70
77 - Refusal	15	21.13	49	51.04	83	63.36	147	49.33
78 - Parental Refusal	45	63.38	0	0.00	0	0.00	45	15.10
Other	2	2.82	11	11.46	7	5.34	20	6.71

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Massachusetts) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	96	100.00	131	100.00	298	100.00
71 - No One at DU	0	0.00	7	7.20	6	3.32	13	3.55
72 - Respondent Unavailable	5	4.26	21	19.86	11	6.42	37	7.73
73 - Break Off (Partial Interview)	0	0.00	1	0.70	1	0.97	2	0.88
74 - Physically/Mentally Incompetent	4	5.63	3	2.92	9	11.00	16	9.84
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.29	1	0.25
76 - Language Barrier - Other	0	0.00	4	3.21	13	11.16	17	9.69
77 - Refusal	15	27.53	49	55.25	83	62.56	147	59.82
78 - Parental Refusal	45	61.36	0	0.00	0	0.00	45	3.43
Other	2	1.22	11	10.87	7	4.27	20	4.80

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Michigan) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	267	100.00	277	100.00	407	100.00	951	100.00
71 - No One at DU	8	3.00	16	5.78	16	3.93	40	4.21
72 - Respondent Unavailable	13	4.87	25	9.03	14	3.44	52	5.47
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	9	3.37	12	4.33	24	5.90	45	4.73
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.37	8	2.89	15	3.69	24	2.52
77 - Refusal	46	17.23	196	70.76	334	82.06	576	60.57
78 - Parental Refusal	178	66.67	0	0.00	0	0.00	178	18.72
Other	12	4.49	20	7.22	4	0.98	36	3.79

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Michigan) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	267	100.00	277	100.00	407	100.00	951	100.00
71 - No One at DU	8	2.66	16	4.90	16	3.01	40	3.17
72 - Respondent Unavailable	13	5.06	25	7.70	14	2.76	52	3.42
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	9	3.10	12	4.48	24	7.37	45	6.77
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.37	8	3.03	15	5.12	24	4.56
77 - Refusal	46	16.70	196	72.62	334	80.67	576	75.18
78 - Parental Refusal	178	67.44	0	0.00	0	0.00	178	4.95
Other	12	4.68	20	7.27	4	1.07	36	1.95

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Minnesota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	39	100.00	67	100.00	84	100.00	190	100.00
71 - No One at DU	0	0.00	5	7.46	4	4.76	9	4.74
72 - Respondent Unavailable	2	5.13	4	5.97	2	2.38	8	4.21
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	2.56	0	0.00	0	0.00	1	0.53
75 - Language Barrier - Spanish	0	0.00	1	1.49	4	4.76	5	2.63
76 - Language Barrier - Other	1	2.56	1	1.49	8	9.52	10	5.26
77 - Refusal	5	12.82	48	71.64	62	73.81	115	60.53
78 - Parental Refusal	30	76.92	0	0.00	0	0.00	30	15.79
Other	0	0.00	8	11.94	4	4.76	12	6.32

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Minnesota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	39	100.00	67	100.00	84	100.00	190	100.00
71 - No One at DU	0	0.00	5	5.96	4	3.21	9	3.43
72 - Respondent Unavailable	2	3.68	4	4.19	2	1.99	8	2.39
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.93	0	0.00	0	0.00	1	0.10
75 - Language Barrier - Spanish	0	0.00	1	0.99	4	2.04	5	1.78
76 - Language Barrier - Other	1	4.60	1	2.19	8	14.56	10	12.28
77 - Refusal	5	12.29	48	77.44	62	75.55	115	72.41
78 - Parental Refusal	30	77.51	0	0.00	0	0.00	30	4.17
Other	0	0.00	8	9.24	4	2.65	12	3.43

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Mississippi) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	41	100.00	87	100.00	199	100.00
71 - No One at DU	7	9.86	12	29.27	10	11.49	29	14.57
72 - Respondent Unavailable	2	2.82	4	9.76	9	10.34	15	7.54
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	5.63	1	2.44	10	11.49	15	7.54
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	1.15	1	0.50
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	15.49	15	36.59	52	59.77	78	39.20
78 - Parental Refusal	37	52.11	0	0.00	0	0.00	37	18.59
Other	10	14.08	9	21.95	5	5.75	24	12.06

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Mississippi) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	71	100.00	41	100.00	87	100.00	199	100.00
71 - No One at DU	7	10.44	12	30.60	10	9.31	29	11.15
72 - Respondent Unavailable	2	3.19	4	12.03	9	9.79	15	9.32
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	5.41	1	0.59	10	14.51	15	12.47
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.69	1	0.57
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	19.04	15	35.45	52	60.68	78	54.51
78 - Parental Refusal	37	45.98	0	0.00	0	0.00	37	4.55
Other	10	15.94	9	21.33	5	5.03	24	7.43

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Missouri) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	55	100.00	66	100.00	113	100.00	234	100.00
71 - No One at DU	5	9.09	6	9.09	7	6.19	18	7.69
72 - Respondent Unavailable	5	9.09	2	3.03	6	5.31	13	5.56
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	12.73	6	9.09	8	7.08	21	8.97
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	5	9.09	48	72.73	92	81.42	145	61.97
78 - Parental Refusal	32	58.18	0	0.00	0	0.00	32	13.68
Other	1	1.82	4	6.06	0	0.00	5	2.14

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Missouri) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	55	100.00	66	100.00	113	100.00	234	100.00
71 - No One at DU	5	7.32	6	7.35	7	6.41	18	6.54
72 - Respondent Unavailable	5	9.19	2	2.31	6	4.29	13	4.38
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	15.94	6	10.40	8	9.21	21	9.68
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	5	9.89	48	71.23	92	80.08	145	75.48
78 - Parental Refusal	32	54.50	0	0.00	0	0.00	32	2.95
Other	1	3.16	4	8.71	0	0.00	5	0.96

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Montana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	72	100.00	71	100.00	90	100.00	233	100.00
71 - No One at DU	0	0.00	1	1.41	3	3.33	4	1.72
72 - Respondent Unavailable	1	1.39	4	5.63	2	2.22	7	3.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	5.56	1	1.41	3	3.33	8	3.43
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	15.28	59	83.10	81	90.00	151	64.81
78 - Parental Refusal	56	77.78	0	0.00	0	0.00	56	24.03
Other	0	0.00	6	8.45	1	1.11	7	3.00

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Montana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	72	100.00	71	100.00	90	100.00	233	100.00
71 - No One at DU	0	0.00	1	2.52	3	3.06	4	2.75
72 - Respondent Unavailable	1	1.39	4	5.39	2	2.74	7	2.99
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	7.12	1	1.93	3	4.75	8	4.56
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	11	14.22	59	81.91	81	88.64	151	81.81
78 - Parental Refusal	56	77.27	0	0.00	0	0.00	56	6.15
Other	0	0.00	6	8.25	1	0.81	7	1.74

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Nebraska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	68	100.00	118	100.00	230	100.00
71 - No One at DU	2	4.55	10	14.71	5	4.24	17	7.39
72 - Respondent Unavailable	5	11.36	16	23.53	8	6.78	29	12.61
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	4.55	1	1.47	6	5.08	9	3.91
75 - Language Barrier - Spanish	0	0.00	1	1.47	0	0.00	1	0.43
76 - Language Barrier - Other	0	0.00	5	7.35	1	0.85	6	2.61
77 - Refusal	5	11.36	34	50.00	96	81.36	135	58.70
78 - Parental Refusal	30	68.18	0	0.00	0	0.00	30	13.04
Other	0	0.00	1	1.47	2	1.69	3	1.30

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Nebraska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	68	100.00	118	100.00	230	100.00
71 - No One at DU	2	3.89	10	13.73	5	4.79	17	5.40
72 - Respondent Unavailable	5	11.51	16	26.69	8	8.05	29	9.57
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	3.96	1	1.06	6	5.98	9	5.53
75 - Language Barrier - Spanish	0	0.00	1	2.64	0	0.00	1	0.19
76 - Language Barrier - Other	0	0.00	5	8.24	1	0.82	6	1.32
77 - Refusal	5	10.05	34	45.45	96	78.63	135	72.96
78 - Parental Refusal	30	70.58	0	0.00	0	0.00	30	3.35
Other	0	0.00	1	2.20	2	1.73	3	1.68

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Nevada) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	43	100.00	79	100.00	109	100.00	231	100.00
71 - No One at DU	2	4.65	4	5.06	5	4.59	11	4.76
72 - Respondent Unavailable	1	2.33	2	2.53	4	3.67	7	3.03
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	2.33	2	2.53	5	4.59	8	3.46
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.92	1	0.43
76 - Language Barrier - Other	0	0.00	1	1.27	2	1.83	3	1.30
77 - Refusal	15	34.88	63	79.75	89	81.65	167	72.29
78 - Parental Refusal	22	51.16	0	0.00	0	0.00	22	9.52
Other	2	4.65	7	8.86	3	2.75	12	5.19

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Nevada) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	43	100.00	79	100.00	109	100.00	231	100.00
71 - No One at DU	2	6.68	4	5.26	5	5.02	11	5.13
72 - Respondent Unavailable	1	3.73	2	3.32	4	2.53	7	2.68
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.92	2	3.19	5	3.81	8	3.65
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.50	1	0.42
76 - Language Barrier - Other	0	0.00	1	1.00	2	1.51	3	1.38
77 - Refusal	15	40.14	63	78.77	89	85.05	167	82.04
78 - Parental Refusal	22	45.16	0	0.00	0	0.00	22	2.37
Other	2	2.37	7	8.47	3	1.58	12	2.33

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (New Hampshire) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	100	100.00	93	100.00	116	100.00	309	100.00
71 - No One at DU	1	1.00	3	3.23	1	0.86	5	1.62
72 - Respondent Unavailable	1	1.00	9	9.68	2	1.72	12	3.88
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	10	10.00	3	3.23	6	5.17	19	6.15
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	1.00	0	0.00	5	4.31	6	1.94
77 - Refusal	10	10.00	75	80.65	102	87.93	187	60.52
78 - Parental Refusal	77	77.00	0	0.00	0	0.00	77	24.92
Other	0	0.00	3	3.23	0	0.00	3	0.97

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (New Hampshire) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	100	100.00	93	100.00	116	100.00	309	100.00
71 - No One at DU	1	0.84	3	2.88	1	0.72	5	0.94
72 - Respondent Unavailable	1	1.29	9	8.11	2	1.13	12	1.83
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	10	8.35	3	2.76	6	5.53	19	5.50
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	1.27	0	0.00	5	4.95	6	4.15
77 - Refusal	10	10.21	75	82.74	102	87.68	187	80.59
78 - Parental Refusal	77	78.03	0	0.00	0	0.00	77	6.65
Other	0	0.00	3	3.52	0	0.00	3	0.35

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (New Jersey) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	58	100.00	86	100.00	113	100.00	257	100.00
71 - No One at DU	0	0.00	5	5.81	2	1.77	7	2.72
72 - Respondent Unavailable	0	0.00	2	2.33	1	0.88	3	1.17
73 - Break Off (Partial Interview)	0	0.00	2	2.33	1	0.88	3	1.17
74 - Physically/Mentally Incompetent	3	5.17	3	3.49	8	7.08	14	5.45
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.88	1	0.39
76 - Language Barrier - Other	0	0.00	0	0.00	5	4.42	5	1.95
77 - Refusal	7	12.07	69	80.23	92	81.42	168	65.37
78 - Parental Refusal	48	82.76	0	0.00	0	0.00	48	18.68
Other	0	0.00	5	5.81	3	2.65	8	3.11

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (New Jersey) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	58	100.00	86	100.00	113	100.00	257	100.00
71 - No One at DU	0	0.00	5	5.84	2	1.57	7	1.91
72 - Respondent Unavailable	0	0.00	2	1.73	1	0.37	3	0.49
73 - Break Off (Partial Interview)	0	0.00	2	1.87	1	0.50	3	0.61
74 - Physically/Mentally Incompetent	3	6.75	3	4.23	8	9.07	14	8.43
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.33	1	0.27
76 - Language Barrier - Other	0	0.00	0	0.00	5	6.23	5	5.21
77 - Refusal	7	10.39	69	80.70	92	80.62	168	76.28
78 - Parental Refusal	48	82.86	0	0.00	0	0.00	48	5.13
Other	0	0.00	5	5.64	3	1.30	8	1.67

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (New Mexico) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	42	100.00	66	100.00	114	100.00	222	100.00
71 - No One at DU	1	2.38	3	4.55	5	4.39	9	4.05
72 - Respondent Unavailable	0	0.00	2	3.03	2	1.75	4	1.80
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	4.76	2	3.03	5	4.39	9	4.05
75 - Language Barrier - Spanish	0	0.00	1	1.52	0	0.00	1	0.45
76 - Language Barrier - Other	0	0.00	1	1.52	4	3.51	5	2.25
77 - Refusal	12	28.57	53	80.30	94	82.46	159	71.62
78 - Parental Refusal	25	59.52	0	0.00	0	0.00	25	11.26
Other	2	4.76	4	6.06	4	3.51	10	4.50

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (New Mexico) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	42	100.00	66	100.00	114	100.00	222	100.00
71 - No One at DU	1	1.66	3	2.94	5	4.06	9	3.83
72 - Respondent Unavailable	0	0.00	2	2.70	2	0.93	4	1.05
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	3.40	2	1.75	5	6.29	9	5.71
75 - Language Barrier - Spanish	0	0.00	1	1.25	0	0.00	1	0.12
76 - Language Barrier - Other	0	0.00	1	3.10	4	7.96	5	7.09
77 - Refusal	12	27.06	53	82.75	94	76.21	159	74.33
78 - Parental Refusal	25	63.66	0	0.00	0	0.00	25	3.24
Other	2	4.22	4	5.51	4	4.55	10	4.62

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (New York) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	371	100.00	512	100.00	704	100.00	1,587	100.00
71 - No One at DU	13	3.50	50	9.77	59	8.38	122	7.69
72 - Respondent Unavailable	33	8.89	87	16.99	84	11.93	204	12.85
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.14	1	0.06
74 - Physically/Mentally Incompetent	22	5.93	10	1.95	30	4.26	62	3.91
75 - Language Barrier - Spanish	0	0.00	1	0.20	0	0.00	1	0.06
76 - Language Barrier - Other	5	1.35	9	1.76	41	5.82	55	3.47
77 - Refusal	50	13.48	319	62.30	470	66.76	839	52.87
78 - Parental Refusal	242	65.23	0	0.00	0	0.00	242	15.25
Other	6	1.62	36	7.03	19	2.70	61	3.84

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (New York) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	371	100.00	512	100.00	704	100.00	1,587	100.00
71 - No One at DU	13	3.23	50	8.83	59	6.59	122	6.63
72 - Respondent Unavailable	33	8.72	87	16.79	84	11.81	204	12.16
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.24	1	0.20
74 - Physically/Mentally Incompetent	22	6.52	10	1.51	30	5.21	62	4.89
75 - Language Barrier - Spanish	0	0.00	1	0.13	0	0.00	1	0.01
76 - Language Barrier - Other	5	1.39	9	2.17	41	8.41	55	7.31
77 - Refusal	50	13.81	319	62.38	470	64.96	839	61.58
78 - Parental Refusal	242	64.33	0	0.00	0	0.00	242	3.90
Other	6	2.00	36	8.19	19	2.77	61	3.31

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (North Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	56	100.00	45	100.00	99	100.00	200	100.00
71 - No One at DU	4	7.14	2	4.44	2	2.02	8	4.00
72 - Respondent Unavailable	4	7.14	2	4.44	8	8.08	14	7.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	10.71	2	4.44	12	12.12	20	10.00
75 - Language Barrier - Spanish	0	0.00	1	2.22	2	2.02	3	1.50
76 - Language Barrier - Other	1	1.79	1	2.22	0	0.00	2	1.00
77 - Refusal	3	5.36	35	77.78	72	72.73	110	55.00
78 - Parental Refusal	36	64.29	0	0.00	0	0.00	36	18.00
Other	2	3.57	2	4.44	3	3.03	7	3.50

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (North Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	56	100.00	45	100.00	99	100.00	200	100.00
71 - No One at DU	4	6.99	2	3.26	2	3.02	8	3.30
72 - Respondent Unavailable	4	6.89	2	4.80	8	5.27	14	5.35
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	9.77	2	3.33	12	15.20	20	14.08
75 - Language Barrier - Spanish	0	0.00	1	2.93	2	1.72	3	1.68
76 - Language Barrier - Other	1	2.07	1	1.13	0	0.00	2	0.21
77 - Refusal	3	6.09	35	81.47	72	72.07	110	68.22
78 - Parental Refusal	36	63.88	0	0.00	0	0.00	36	4.30
Other	2	4.31	2	3.08	3	2.73	7	2.86

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (North Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	62	100.00	71	100.00	128	100.00	261	100.00
71 - No One at DU	0	0.00	0	0.00	1	0.78	1	0.38
72 - Respondent Unavailable	1	1.61	5	7.04	6	4.69	12	4.60
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.61	3	4.23	2	1.56	6	2.30
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.78	1	0.38
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.78	1	0.38
77 - Refusal	27	43.55	61	85.92	116	90.63	204	78.16
78 - Parental Refusal	33	53.23	0	0.00	0	0.00	33	12.64
Other	0	0.00	2	2.82	1	0.78	3	1.15

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (North Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	62	100.00	71	100.00	128	100.00	261	100.00
71 - No One at DU	0	0.00	0	0.00	1	0.34	1	0.28
72 - Respondent Unavailable	1	1.13	5	6.37	6	3.65	12	3.85
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.34	3	3.17	2	1.10	6	1.36
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.50	1	0.41
76 - Language Barrier - Other	0	0.00	0	0.00	1	1.33	1	1.10
77 - Refusal	27	40.21	61	87.86	116	92.58	204	89.27
78 - Parental Refusal	33	57.32	0	0.00	0	0.00	33	3.00
Other	0	0.00	2	2.60	1	0.50	3	0.73

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Ohio) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	331	100.00	327	100.00	482	100.00	1,140	100.00
71 - No One at DU	18	5.44	31	9.48	23	4.77	72	6.32
72 - Respondent Unavailable	20	6.04	30	9.17	31	6.43	81	7.11
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	20	6.04	12	3.67	35	7.26	67	5.88
75 - Language Barrier - Spanish	1	0.30	4	1.22	3	0.62	8	0.70
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.83	4	0.35
77 - Refusal	59	17.82	221	67.58	376	78.01	656	57.54
78 - Parental Refusal	199	60.12	0	0.00	0	0.00	199	17.46
Other	14	4.23	29	8.87	10	2.07	53	4.65

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Ohio) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	331	100.00	327	100.00	482	100.00	1,140	100.00
71 - No One at DU	18	6.40	31	9.10	23	4.99	72	5.52
72 - Respondent Unavailable	20	6.21	30	9.35	31	6.09	81	6.43
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	20	7.22	12	3.51	35	9.54	67	8.75
75 - Language Barrier - Spanish	1	0.29	4	1.23	3	0.47	8	0.53
76 - Language Barrier - Other	0	0.00	0	0.00	4	0.68	4	0.56
77 - Refusal	59	17.15	221	67.61	376	76.34	656	71.07
78 - Parental Refusal	199	58.55	0	0.00	0	0.00	199	4.33
Other	14	4.18	29	9.20	10	1.89	53	2.81

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Oklahoma) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	82	100.00	86	100.00	113	100.00	281	100.00
71 - No One at DU	1	1.22	2	2.33	1	0.88	4	1.42
72 - Respondent Unavailable	4	4.88	2	2.33	5	4.42	11	3.91
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	6.10	2	2.33	8	7.08	15	5.34
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	13	15.85	71	82.56	96	84.96	180	64.06
78 - Parental Refusal	58	70.73	0	0.00	0	0.00	58	20.64
Other	1	1.22	9	10.47	3	2.65	13	4.63

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Oklahoma) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	82	100.00	86	100.00	113	100.00	281	100.00
71 - No One at DU	1	0.87	2	1.77	1	1.06	4	1.13
72 - Respondent Unavailable	4	6.00	2	2.59	5	4.37	11	4.29
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	4.95	2	3.25	8	7.02	15	6.42
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	13	16.11	71	83.40	96	85.47	180	79.79
78 - Parental Refusal	58	71.00	0	0.00	0	0.00	58	5.57
Other	1	1.07	9	8.99	3	2.07	13	2.81

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Oregon) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	41	100.00	89	100.00	112	100.00	242	100.00
71 - No One at DU	0	0.00	5	5.62	3	2.68	8	3.31
72 - Respondent Unavailable	5	12.20	8	8.99	14	12.50	27	11.16
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	9.76	2	2.25	5	4.46	11	4.55
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	1.79	2	0.83
76 - Language Barrier - Other	0	0.00	0	0.00	3	2.68	3	1.24
77 - Refusal	13	31.71	55	61.80	82	73.21	150	61.98
78 - Parental Refusal	17	41.46	0	0.00	0	0.00	17	7.02
Other	2	4.88	19	21.35	3	2.68	24	9.92

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Oregon) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	41	100.00	89	100.00	112	100.00	242	100.00
71 - No One at DU	0	0.00	5	6.75	3	2.91	8	3.17
72 - Respondent Unavailable	5	14.39	8	8.23	14	11.34	27	11.15
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	9.23	2	2.35	5	6.75	11	6.41
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	1.73	2	1.48
76 - Language Barrier - Other	0	0.00	0	0.00	3	5.30	3	4.52
77 - Refusal	13	34.76	55	60.32	82	70.09	150	67.53
78 - Parental Refusal	17	37.82	0	0.00	0	0.00	17	1.67
Other	2	3.79	19	22.35	3	1.87	24	4.06

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Pennsylvania) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	256	100.00	318	100.00	551	100.00	1,125	100.00
71 - No One at DU	9	3.52	25	7.86	29	5.26	63	5.60
72 - Respondent Unavailable	13	5.08	34	10.69	14	2.54	61	5.42
73 - Break Off (Partial Interview)	0	0.00	1	0.31	0	0.00	1	0.09
74 - Physically/Mentally Incompetent	17	6.64	10	3.14	38	6.90	65	5.78
75 - Language Barrier - Spanish	0	0.00	1	0.31	5	0.91	6	0.53
76 - Language Barrier - Other	0	0.00	3	0.94	10	1.81	13	1.16
77 - Refusal	59	23.05	229	72.01	447	81.13	735	65.33
78 - Parental Refusal	149	58.20	0	0.00	0	0.00	149	13.24
Other	9	3.52	15	4.72	8	1.45	32	2.84

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Pennsylvania) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	256	100.00	318	100.00	551	100.00	1,125	100.00
71 - No One at DU	9	2.73	25	7.82	29	4.57	63	4.76
72 - Respondent Unavailable	13	5.47	34	11.13	14	1.86	61	2.88
73 - Break Off (Partial Interview)	0	0.00	1	0.16	0	0.00	1	0.01
74 - Physically/Mentally Incompetent	17	5.77	10	2.44	38	8.92	65	8.17
75 - Language Barrier - Spanish	0	0.00	1	0.30	5	1.04	6	0.92
76 - Language Barrier - Other	0	0.00	3	1.37	10	3.00	13	2.70
77 - Refusal	59	21.95	229	71.79	447	79.18	735	75.49
78 - Parental Refusal	149	60.61	0	0.00	0	0.00	149	3.21
Other	9	3.47	15	4.99	8	1.43	32	1.86

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Rhode Island) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	62	100.00	102	100.00	208	100.00
71 - No One at DU	1	2.27	4	6.45	2	1.96	7	3.37
72 - Respondent Unavailable	2	4.55	6	9.68	4	3.92	12	5.77
73 - Break Off (Partial Interview)	0	0.00	1	1.61	0	0.00	1	0.48
74 - Physically/Mentally Incompetent	1	2.27	3	4.84	6	5.88	10	4.81
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	2.27	0	0.00	3	2.94	4	1.92
77 - Refusal	9	20.45	43	69.35	85	83.33	137	65.87
78 - Parental Refusal	28	63.64	0	0.00	0	0.00	28	13.46
Other	2	4.55	5	8.06	2	1.96	9	4.33

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Rhode Island) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	62	100.00	102	100.00	208	100.00
71 - No One at DU	1	1.56	4	5.97	2	2.29	7	2.63
72 - Respondent Unavailable	2	4.58	6	9.12	4	4.05	12	4.59
73 - Break Off (Partial Interview)	0	0.00	1	1.86	0	0.00	1	0.19
74 - Physically/Mentally Incompetent	1	1.67	3	5.35	6	7.19	10	6.71
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.51	0	0.00	3	2.88	4	2.47
77 - Refusal	9	22.81	43	70.04	85	81.66	137	77.44
78 - Parental Refusal	28	66.94	0	0.00	0	0.00	28	3.45
Other	2	1.92	5	7.66	2	1.93	9	2.52

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (South Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	68	100.00	54	100.00	111	100.00	233	100.00
71 - No One at DU	2	2.94	10	18.52	5	4.50	17	7.30
72 - Respondent Unavailable	3	4.41	6	11.11	17	15.32	26	11.16
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.90	1	0.43
74 - Physically/Mentally Incompetent	4	5.88	5	9.26	9	8.11	18	7.73
75 - Language Barrier - Spanish	0	0.00	1	1.85	3	2.70	4	1.72
76 - Language Barrier - Other	0	0.00	0	0.00	1	0.90	1	0.43
77 - Refusal	1	1.47	27	50.00	71	63.96	99	42.49
78 - Parental Refusal	53	77.94	0	0.00	0	0.00	53	22.75
Other	5	7.35	5	9.26	4	3.60	14	6.01

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (South Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	68	100.00	54	100.00	111	100.00	233	100.00
71 - No One at DU	2	2.81	10	17.22	5	4.20	17	5.17
72 - Respondent Unavailable	3	5.43	6	10.20	17	13.33	26	12.52
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.97	1	0.82
74 - Physically/Mentally Incompetent	4	5.59	5	10.96	9	8.12	18	8.18
75 - Language Barrier - Spanish	0	0.00	1	1.03	3	1.33	4	1.21
76 - Language Barrier - Other	0	0.00	0	0.00	1	2.63	1	2.23
77 - Refusal	1	1.44	27	50.87	71	66.51	99	60.67
78 - Parental Refusal	53	78.86	0	0.00	0	0.00	53	5.52
Other	5	5.86	5	9.73	4	2.91	14	3.67

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (South Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	51	100.00	71	100.00	113	100.00	235	100.00
71 - No One at DU	5	9.80	8	11.27	10	8.85	23	9.79
72 - Respondent Unavailable	1	1.96	4	5.63	3	2.65	8	3.40
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	1.96	7	9.86	6	5.31	14	5.96
75 - Language Barrier - Spanish	0	0.00	3	4.23	4	3.54	7	2.98
76 - Language Barrier - Other	0	0.00	0	0.00	3	2.65	3	1.28
77 - Refusal	11	21.57	37	52.11	82	72.57	130	55.32
78 - Parental Refusal	29	56.86	0	0.00	0	0.00	29	12.34
Other	4	7.84	12	16.90	5	4.42	21	8.94

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (South Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	51	100.00	71	100.00	113	100.00	235	100.00
71 - No One at DU	5	13.73	8	8.52	10	6.52	23	7.13
72 - Respondent Unavailable	1	1.46	4	6.76	3	1.54	8	2.01
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	1	2.61	7	8.03	6	6.95	14	6.79
75 - Language Barrier - Spanish	0	0.00	3	3.82	4	3.23	7	3.09
76 - Language Barrier - Other	0	0.00	0	0.00	3	1.96	3	1.66
77 - Refusal	11	21.03	37	54.34	82	75.59	130	70.41
78 - Parental Refusal	29	53.48	0	0.00	0	0.00	29	3.18
Other	4	7.69	12	18.52	5	4.22	21	5.73

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Tennessee) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	38	100.00	67	100.00	73	100.00	178	100.00
71 - No One at DU	13	34.21	15	22.39	4	5.48	32	17.98
72 - Respondent Unavailable	0	0.00	5	7.46	5	6.85	10	5.62
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	5.26	2	2.99	2	2.74	6	3.37
75 - Language Barrier - Spanish	0	0.00	3	4.48	6	8.22	9	5.06
76 - Language Barrier - Other	0	0.00	0	0.00	2	2.74	2	1.12
77 - Refusal	3	7.89	36	53.73	54	73.97	93	52.25
78 - Parental Refusal	19	50.00	0	0.00	0	0.00	19	10.67
Other	1	2.63	6	8.96	0	0.00	7	3.93

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Tennessee) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	38	100.00	67	100.00	73	100.00	178	100.00
71 - No One at DU	13	32.01	15	22.06	4	4.58	32	8.42
72 - Respondent Unavailable	0	0.00	5	6.70	5	4.77	10	4.67
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	2.76	2	3.63	2	4.03	6	3.90
75 - Language Barrier - Spanish	0	0.00	3	5.43	6	5.94	9	5.48
76 - Language Barrier - Other	0	0.00	0	0.00	2	7.20	2	5.90
77 - Refusal	3	5.90	36	54.32	54	73.48	93	66.74
78 - Parental Refusal	19	56.70	0	0.00	0	0.00	19	3.83
Other	1	2.63	6	7.85	0	0.00	7	1.07

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Texas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	226	100.00	288	100.00	473	100.00	987	100.00
71 - No One at DU	19	8.41	52	18.06	68	14.38	139	14.08
72 - Respondent Unavailable	20	8.85	45	15.63	48	10.15	113	11.45
73 - Break Off (Partial Interview)	0	0.00	1	0.35	2	0.42	3	0.30
74 - Physically/Mentally Incompetent	15	6.64	12	4.17	21	4.44	48	4.86
75 - Language Barrier - Spanish	0	0.00	5	1.74	7	1.48	12	1.22
76 - Language Barrier - Other	0	0.00	1	0.35	11	2.33	12	1.22
77 - Refusal	30	13.27	145	50.35	305	64.48	480	48.63
78 - Parental Refusal	126	55.75	0	0.00	0	0.00	126	12.77
Other	16	7.08	27	9.38	11	2.33	54	5.47

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Texas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	226	100.00	288	100.00	473	100.00	987	100.00
71 - No One at DU	19	8.51	52	18.01	68	12.64	139	12.93
72 - Respondent Unavailable	20	8.39	45	14.91	48	9.94	113	10.35
73 - Break Off (Partial Interview)	0	0.00	1	0.25	2	0.47	3	0.41
74 - Physically/Mentally Incompetent	15	6.40	12	4.48	21	5.77	48	5.68
75 - Language Barrier - Spanish	0	0.00	5	1.15	7	1.58	12	1.43
76 - Language Barrier - Other	0	0.00	1	0.38	11	2.91	12	2.46
77 - Refusal	30	13.26	145	52.02	305	64.35	480	59.81
78 - Parental Refusal	126	56.66	0	0.00	0	0.00	126	3.63
Other	16	6.79	27	8.80	11	2.35	54	3.30

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Utah) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	32	100.00	74	100.00	67	100.00	173	100.00
71 - No One at DU	0	0.00	5	6.76	7	10.45	12	6.94
72 - Respondent Unavailable	6	18.75	12	16.22	10	14.93	28	16.18
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	1.49	1	0.58
74 - Physically/Mentally Incompetent	2	6.25	2	2.70	2	2.99	6	3.47
75 - Language Barrier - Spanish	0	0.00	3	4.05	0	0.00	3	1.73
76 - Language Barrier - Other	0	0.00	1	1.35	3	4.48	4	2.31
77 - Refusal	7	21.88	46	62.16	44	65.67	97	56.07
78 - Parental Refusal	16	50.00	0	0.00	0	0.00	16	9.25
Other	1	3.13	5	6.76	0	0.00	6	3.47

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Utah) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	32	100.00	74	100.00	67	100.00	173	100.00
71 - No One at DU	0	0.00	5	5.45	7	6.56	12	5.88
72 - Respondent Unavailable	6	23.00	12	17.04	10	14.86	28	15.85
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	1.05	1	0.78
74 - Physically/Mentally Incompetent	2	6.76	2	3.12	2	3.85	6	3.92
75 - Language Barrier - Spanish	0	0.00	3	3.86	0	0.00	3	0.73
76 - Language Barrier - Other	0	0.00	1	1.57	3	6.62	4	5.19
77 - Refusal	7	15.69	46	60.90	44	67.06	97	62.25
78 - Parental Refusal	16	51.49	0	0.00	0	0.00	16	3.65
Other	1	3.06	5	8.06	0	0.00	6	1.75

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Vermont) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	65	100.00	78	100.00	108	100.00	251	100.00
71 - No One at DU	4	6.15	7	8.97	5	4.63	16	6.37
72 - Respondent Unavailable	8	12.31	7	8.97	10	9.26	25	9.96
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	4.62	3	3.85	4	3.70	10	3.98
75 - Language Barrier - Spanish	1	1.54	0	0.00	0	0.00	1	0.40
76 - Language Barrier - Other	0	0.00	0	0.00	4	3.70	4	1.59
77 - Refusal	19	29.23	50	64.10	80	74.07	149	59.36
78 - Parental Refusal	28	43.08	0	0.00	0	0.00	28	11.16
Other	2	3.08	11	14.10	5	4.63	18	7.17

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Vermont) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	65	100.00	78	100.00	108	100.00	251	100.00
71 - No One at DU	4	5.05	7	7.78	5	4.47	16	4.83
72 - Respondent Unavailable	8	12.30	7	13.16	10	8.66	25	9.32
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	6.07	3	3.58	4	3.96	10	4.05
75 - Language Barrier - Spanish	1	0.97	0	0.00	0	0.00	1	0.06
76 - Language Barrier - Other	0	0.00	0	0.00	4	6.41	4	5.41
77 - Refusal	19	25.73	50	60.93	80	71.77	149	67.96
78 - Parental Refusal	28	48.43	0	0.00	0	0.00	28	2.90
Other	2	1.45	11	14.55	5	4.73	18	5.49

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	51	100.00	46	100.00	104	100.00	201	100.00
71 - No One at DU	0	0.00	1	2.17	4	3.85	5	2.49
72 - Respondent Unavailable	7	13.73	7	15.22	8	7.69	22	10.95
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	11.76	0	0.00	4	3.85	10	4.98
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	2.17	6	5.77	7	3.48
77 - Refusal	9	17.65	36	78.26	81	77.88	126	62.69
78 - Parental Refusal	27	52.94	0	0.00	0	0.00	27	13.43
Other	2	3.92	1	2.17	1	0.96	4	1.99

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	51	100.00	46	100.00	104	100.00	201	100.00
71 - No One at DU	0	0.00	1	2.20	4	3.00	5	2.75
72 - Respondent Unavailable	7	13.81	7	20.80	8	6.11	22	7.82
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	6	12.78	0	0.00	4	5.19	10	5.21
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	1	2.04	6	6.54	7	5.77
77 - Refusal	9	15.23	36	71.66	81	78.50	126	74.14
78 - Parental Refusal	27	52.99	0	0.00	0	0.00	27	3.17
Other	2	5.19	1	3.30	1	0.65	4	1.14

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Washington) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	67	100.00	96	100.00	127	100.00	290	100.00
71 - No One at DU	2	2.99	5	5.21	5	3.94	12	4.14
72 - Respondent Unavailable	5	7.46	14	14.58	10	7.87	29	10.00
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	2.99	2	2.08	6	4.72	10	3.45
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.79	1	0.34
76 - Language Barrier - Other	2	2.99	1	1.04	9	7.09	12	4.14
77 - Refusal	13	19.40	63	65.63	93	73.23	169	58.28
78 - Parental Refusal	42	62.69	0	0.00	0	0.00	42	14.48
Other	1	1.49	11	11.46	3	2.36	15	5.17

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Washington) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	67	100.00	96	100.00	127	100.00	290	100.00
71 - No One at DU	2	2.03	5	5.49	5	5.21	12	5.07
72 - Respondent Unavailable	5	8.69	14	14.19	10	9.77	29	10.18
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	2.74	2	4.49	6	6.64	10	6.20
75 - Language Barrier - Spanish	0	0.00	0	0.00	1	0.99	1	0.83
76 - Language Barrier - Other	2	4.29	1	1.39	9	9.62	12	8.46
77 - Refusal	13	21.94	63	62.72	93	64.24	169	61.78
78 - Parental Refusal	42	59.12	0	0.00	0	0.00	42	3.22
Other	1	1.20	11	11.72	3	3.52	15	4.26

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (West Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	81	100.00	116	100.00	241	100.00
71 - No One at DU	7	15.91	17	20.99	11	9.48	35	14.52
72 - Respondent Unavailable	3	6.82	15	18.52	3	2.59	21	8.71
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	6.82	2	2.47	12	10.34	17	7.05
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	12	27.27	45	55.56	89	76.72	146	60.58
78 - Parental Refusal	17	38.64	0	0.00	0	0.00	17	7.05
Other	2	4.55	2	2.47	1	0.86	5	2.07

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (West Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	81	100.00	116	100.00	241	100.00
71 - No One at DU	7	14.28	17	19.35	11	7.55	35	8.78
72 - Respondent Unavailable	3	8.11	15	20.12	3	3.06	21	4.64
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	3	5.36	2	2.36	12	12.47	17	11.36
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	0	0.00	0	0.00	0	0.00	0	0.00
77 - Refusal	12	29.13	45	55.54	89	76.11	146	72.56
78 - Parental Refusal	17	37.76	0	0.00	0	0.00	17	1.51
Other	2	5.36	2	2.63	1	0.82	5	1.15

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Wisconsin) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	39	100.00	94	100.00	90	100.00	223	100.00
71 - No One at DU	1	2.56	6	6.38	4	4.44	11	4.93
72 - Respondent Unavailable	0	0.00	10	10.64	4	4.44	14	6.28
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	5.13	2	2.13	7	7.78	11	4.93
75 - Language Barrier - Spanish	0	0.00	2	2.13	4	4.44	6	2.69
76 - Language Barrier - Other	0	0.00	0	0.00	2	2.22	2	0.90
77 - Refusal	17	43.59	71	75.53	68	75.56	156	69.96
78 - Parental Refusal	18	46.15	0	0.00	0	0.00	18	8.07
Other	1	2.56	3	3.19	1	1.11	5	2.24

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Wisconsin) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	39	100.00	94	100.00	90	100.00	223	100.00
71 - No One at DU	1	2.14	6	5.68	4	3.54	11	3.72
72 - Respondent Unavailable	0	0.00	10	11.85	4	4.93	14	5.49
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	2	6.98	2	2.33	7	12.78	11	11.28
75 - Language Barrier - Spanish	0	0.00	2	0.91	4	3.20	6	2.77
76 - Language Barrier - Other	0	0.00	0	0.00	2	4.20	2	3.50
77 - Refusal	17	40.71	71	76.75	68	70.72	156	69.91
78 - Parental Refusal	18	48.42	0	0.00	0	0.00	18	2.43
Other	1	1.75	3	2.48	1	0.62	5	0.89

DU = dwelling unit.

Table 7.22 2012 Interview Results, by Age and Incomplete Interview Result (Wyoming) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	73	100.00	103	100.00	220	100.00
71 - No One at DU	2	4.55	9	12.33	6	5.83	17	7.73
72 - Respondent Unavailable	4	9.09	15	20.55	10	9.71	29	13.18
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	11.36	2	2.74	8	7.77	15	6.82
75 - Language Barrier - Spanish	1	2.27	0	0.00	1	0.97	2	0.91
76 - Language Barrier - Other	0	0.00	1	1.37	2	1.94	3	1.36
77 - Refusal	8	18.18	41	56.16	71	68.93	120	54.55
78 - Parental Refusal	20	45.45	0	0.00	0	0.00	20	9.09
Other	4	9.09	5	6.85	5	4.85	14	6.36

DU = dwelling unit.

Table 7.23 2012 Interview Results, by Age and Incomplete Interview Result (Wyoming) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Incomplete Interview Cases	44	100.00	73	100.00	103	100.00	220	100.00
71 - No One at DU	2	4.56	9	12.95	6	4.29	17	5.32
72 - Respondent Unavailable	4	7.51	15	18.70	10	9.99	29	10.86
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	5	11.20	2	1.90	8	9.53	15	8.73
75 - Language Barrier - Spanish	1	2.16	0	0.00	1	0.67	2	0.67
76 - Language Barrier - Other	0	0.00	1	1.07	2	1.99	3	1.77
77 - Refusal	8	16.32	41	57.65	71	69.90	120	65.41
78 - Parental Refusal	20	50.24	0	0.00	0	0.00	20	2.87
Other	4	8.02	5	7.73	5	3.64	14	4.36

DU = dwelling unit.

Table 7.23a 2012 Interview Results, by Age and Race/Ethnicity, and Incomplete Interview Result (Total United States) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Hispanic or Latino								
Incomplete Interview Cases	642	100.00	1,009	100.00	1,087	100.00	2,738	100.00
71 - No One at DU	30	4.31	90	7.75	89	6.71	209	6.69
72 - Respondent Unavailable	52	6.96	138	14.07	134	12.03	324	11.97
73 - Break Off (Partial Interview)	0	0.00	3	0.33	3	0.35	6	0.32
74 - Physically/Mentally Incompetent	52	8.18	32	3.65	41	5.16	125	5.16
75 - Language Barrier - Spanish	7	0.70	38	2.96	64	4.65	109	4.13
76 - Language Barrier - Other	2	0.35	0	0.00	7	1.21	9	0.98
77 - Refusal	125	23.10	619	61.84	693	64.52	1,437	61.22
78 - Parental Refusal	326	49.52	0	0.00	0	0.00	326	3.48
Other	48	6.89	89	9.40	56	5.36	193	6.05
Not Hispanic or Latino Black or African American								
Incomplete Interview Cases	436	100.00	517	100.00	688	100.00	1,641	100.00
71 - No One at DU	24	5.01	50	9.67	52	6.69	126	6.90
72 - Respondent Unavailable	44	10.78	63	11.13	55	8.47	162	8.93
73 - Break Off (Partial Interview)	0	0.00	0	0.00	3	0.49	3	0.40
74 - Physically/Mentally Incompetent	30	9.59	22	5.00	67	11.84	119	10.92
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	4	1.15	6	0.82	21	2.55	31	2.26
77 - Refusal	86	19.15	310	60.81	461	66.37	857	62.35
78 - Parental Refusal	215	47.91	0	0.00	0	0.00	215	3.46
Other	33	6.41	66	12.57	29	3.59	128	4.79

Table 7.23a 2012 Interview Results, by Age and Race/Ethnicity, and Incomplete Interview Result (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino White								
Incomplete Interview Cases	3,120	100.00	3,821	100.00	6,268	100.00	13,209	100.00
71 - No One at DU	119	3.87	310	7.17	286	3.83	715	4.13
72 - Respondent Unavailable	152	4.67	348	8.79	338	4.84	838	5.19
73 - Break Off (Partial Interview)	0	0.00	5	0.12	6	0.16	11	0.15
74 - Physically/Mentally Incompetent	169	6.32	140	3.55	397	8.38	706	7.82
75 - Language Barrier - Spanish	2	0.01	1	0.02	1	0.01	4	0.01
76 - Language Barrier - Other	6	0.29	17	0.54	89	1.72	112	1.53
77 - Refusal	619	18.91	2,722	72.10	5,019	78.76	8,360	74.75
78 - Parental Refusal	1,966	62.53	0	0.00	0	0.00	1,966	3.55
Other	87	3.39	278	7.71	132	2.30	497	2.86
Not Hispanic or Latino American Indian or Alaska Native								
Incomplete Interview Cases	51	100.00	52	100.00	63	100.00	166	100.00
71 - No One at DU	3	3.92	5	4.63	3	0.99	11	1.82
72 - Respondent Unavailable	7	15.64	4	3.70	2	3.27	13	4.44
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	4	6.02	3	1.28	5	4.70	12	4.28
75 - Language Barrier - Spanish	0	0.00	1	1.08	0	0.00	1	0.17
76 - Language Barrier - Other	0	0.00	0	0.00	3	9.67	3	7.28
77 - Refusal	8	13.91	35	82.15	46	66.81	89	64.55
78 - Parental Refusal	27	59.06	0	0.00	0	0.00	27	5.24
Other	2	1.45	4	7.16	4	14.56	10	12.23

Table 7.23a 2012 Interview Results, by Age and Race/Ethnicity, and Incomplete Interview Result (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino								
Native Hawaiian or Other Pacific Islander								
Incomplete Interview Cases	19	100.00	25	100.00	29	100.00	73	100.00
71 - No One at DU	4	9.91	1	2.67	1	0.68	6	1.37
72 - Respondent Unavailable	3	5.88	3	5.54	3	12.76	9	11.30
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	0	0.00	0	0.00	1	0.76	1	0.60
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	2.58	1	9.96	1	2.73	3	3.92
77 - Refusal	2	3.81	13	51.00	22	82.10	37	73.85
78 - Parental Refusal	9	77.82	0	0.00	0	0.00	9	3.09
Other	0	0.00	7	30.83	1	0.98	8	5.87
Not Hispanic or Latino								
Asian								
Incomplete Interview Cases	203	100.00	307	100.00	570	100.00	1,080	100.00
71 - No One at DU	7	1.43	22	4.86	23	3.08	52	3.15
72 - Respondent Unavailable	7	4.60	27	8.37	28	3.79	62	4.25
73 - Break Off (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00
74 - Physically/Mentally Incompetent	7	3.16	9	3.56	13	2.26	29	2.42
75 - Language Barrier - Spanish	0	0.00	0	0.00	2	0.16	2	0.14
76 - Language Barrier - Other	16	7.50	46	14.69	195	45.71	257	40.92
77 - Refusal	32	20.70	178	60.60	297	43.51	507	43.89
78 - Parental Refusal	125	59.73	0	0.00	0	0.00	125	3.09
Other	9	2.87	25	7.93	12	1.49	46	2.14

Table 7.23a 2012 Interview Results, by Age and Race/Ethnicity, and Incomplete Interview Result (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino Multiple Races								
Incomplete Interview Cases	184	100.00	146	100.00	110	100.00	440	100.00
71 - No One at DU	5	3.53	14	10.24	9	6.95	28	6.92
72 - Respondent Unavailable	11	6.24	15	6.10	11	11.74	37	9.60
73 - Break Off (Partial Interview)	0	0.00	0	0.00	1	0.60	1	0.37
74 - Physically/Mentally Incompetent	12	5.62	10	6.06	9	3.69	31	4.52
75 - Language Barrier - Spanish	0	0.00	0	0.00	0	0.00	0	0.00
76 - Language Barrier - Other	1	0.04	1	0.71	2	4.18	4	2.72
77 - Refusal	28	17.69	99	70.32	74	70.35	201	60.16
78 - Parental Refusal	119	62.90	0	0.00	0	0.00	119	12.17
Other	8	3.97	7	6.59	4	2.49	19	3.56

DU = dwelling unit.

Table 7.24 2012 Interview Refusal Reasons, by Age (Total United States) (Unweighted Percentages)

	12-17		18-25		26+		26-34		35-49		50+		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Refusal Cases	3,687	100.00	3,976	100.00	6,612	100.00	1,473	100.00	2,549	100.00	2,590	100.00	14,275	100.00
Parental refusal	2,787	75.59	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2,787	19.52
Nothing in it for me	487	13.21	1,964	49.40	3,234	48.91	736	49.97	1,226	48.10	1,272	49.11	5,685	39.82
No time	176	4.77	1,028	25.86	1,810	27.37	450	30.55	792	31.07	568	21.93	3,014	21.11
Government/surveys too invasive	99	2.69	234	5.89	743	11.24	116	7.88	246	9.65	381	14.71	1,076	7.54
Gatekeeper/household member won't allow participation	99	2.69	534	13.43	302	4.57	78	5.30	113	4.43	111	4.29	935	6.55
Confidentiality or survey legitimacy concerns	22	0.60	138	3.47	343	5.19	61	4.14	123	4.83	159	6.14	503	3.52
House too messy/too ill	1	0.03	5	0.13	76	1.15	7	0.48	11	0.43	58	2.24	82	0.57
Other	15	0.41	66	1.66	95	1.44	22	1.49	36	1.41	37	1.43	176	1.23
Missing	1	0.03	7	0.18	9	0.14	3	0.20	2	0.08	4	0.15	17	0.12

Table 7.25 2012 Interview Refusal Reasons, by Age (Total United States) (Weighted Percentages)

	12-17		18-25		26+		26-34		35-49		50+		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Refusal Cases	3,687	100.00	3,976	100.00	6,612	100.00	1,473	100.00	2,549	100.00	2,590	100.00	14,275	100.00
Parental refusal	2,787	74.89	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00	2,787	4.93
Nothing in it for me	487	13.34	1,964	50.39	3,234	49.97	736	50.50	1,226	50.04	1,272	49.79	5,685	47.60
No time	176	4.23	1,028	23.25	1,810	25.06	450	28.81	792	29.00	568	21.92	3,014	23.52
Government/surveys too invasive	99	2.93	234	5.88	743	11.23	116	7.35	246	9.27	381	13.33	1,076	10.17
Gatekeeper/household member won't allow participation	99	2.58	534	14.61	302	5.05	78	6.28	113	4.93	111	4.77	935	5.80
Confidentiality or survey legitimacy concerns	22	1.49	138	3.54	343	5.84	61	4.74	123	4.91	159	6.64	503	5.33
House too messy/too ill	1	0.03	5	0.31	76	1.31	7	0.47	11	0.39	58	2.04	82	1.13
Other	15	0.49	66	1.89	95	1.45	22	1.67	36	1.36	37	1.43	176	1.43
Missing	1	0.02	7	0.13	9	0.09	3	0.17	2	0.09	4	0.07	17	0.09

Table 7.24 2012 Interview Refusal Reasons, by Age (Alabama) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	40	100.00	73	100.00	160	100.00
Parental refusal	31	65.96	0	0.00	0	0.00	31	19.38
Nothing in it for me	9	19.15	24	60.00	47	64.38	80	50.00
No time	0	0.00	15	37.50	18	24.66	33	20.63
Government/surveys too invasive	1	2.13	0	0.00	4	5.48	5	3.13
Gatekeeper/household member won't allow participation	5	10.64	1	2.50	3	4.11	9	5.63
Confidentiality or survey legitimacy concerns	1	2.13	0	0.00	1	1.37	2	1.25
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Alabama) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	40	100.00	73	100.00	160	100.00
Parental refusal	31	67.07	0	0.00	0	0.00	31	5.75
Nothing in it for me	9	17.81	24	60.59	47	63.64	80	59.44
No time	0	0.00	15	37.34	18	25.15	33	24.06
Government/surveys too invasive	1	1.28	0	0.00	4	6.83	5	5.76
Gatekeeper/household member won't allow participation	5	12.51	1	2.08	3	3.47	9	4.13
Confidentiality or survey legitimacy concerns	1	1.33	0	0.00	1	0.91	2	0.86
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Alaska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	35	100.00	83	100.00	165	100.00
Parental refusal	30	63.83	0	0.00	0	0.00	30	18.18
Nothing in it for me	9	19.15	20	57.14	52	62.65	81	49.09
No time	4	8.51	7	20.00	15	18.07	26	15.76
Government/surveys too invasive	0	0.00	1	2.86	3	3.61	4	2.42
Gatekeeper/household member won't allow participation	2	4.26	5	14.29	1	1.20	8	4.85
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.20	1	0.61
House too messy/too ill	0	0.00	0	0.00	1	1.20	1	0.61
Other	2	4.26	2	5.71	9	10.84	13	7.88
Missing	0	0.00	0	0.00	1	1.20	1	0.61

Table 7.25 2012 Interview Refusal Reasons, by Age (Alaska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	35	100.00	83	100.00	165	100.00
Parental refusal	30	68.56	0	0.00	0	0.00	30	5.60
Nothing in it for me	9	16.49	20	57.41	52	68.59	81	63.46
No time	4	8.22	7	20.08	15	14.72	26	14.61
Government/surveys too invasive	0	0.00	1	1.90	3	3.47	4	3.06
Gatekeeper/household member won't allow participation	2	3.32	5	14.46	1	2.70	8	3.68
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	0.74	1	0.62
House too messy/too ill	0	0.00	0	0.00	1	0.98	1	0.82
Other	2	3.41	2	6.15	9	8.07	13	7.54
Missing	0	0.00	0	0.00	1	0.73	1	0.61

Table 7.24 2012 Interview Refusal Reasons, by Age (Arizona) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	47	100.00	55	100.00	137	100.00
Parental refusal	15	42.86	0	0.00	0	0.00	15	10.95
Nothing in it for me	10	28.57	27	57.45	28	50.91	65	47.45
No time	6	17.14	10	21.28	16	29.09	32	23.36
Government/surveys too invasive	4	11.43	2	4.26	7	12.73	13	9.49
Gatekeeper/household member won't allow participation	0	0.00	5	10.64	1	1.82	6	4.38
Confidentiality or survey legitimacy concerns	0	0.00	2	4.26	3	5.45	5	3.65
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	1	2.13	0	0.00	1	0.73

Table 7.25 2012 Interview Refusal Reasons, by Age (Arizona) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	47	100.00	55	100.00	137	100.00
Parental refusal	15	40.26	0	0.00	0	0.00	15	2.53
Nothing in it for me	10	25.79	27	62.60	28	49.87	65	50.22
No time	6	16.42	10	16.71	16	31.66	32	28.52
Government/surveys too invasive	4	17.53	2	4.01	7	11.59	13	10.85
Gatekeeper/household member won't allow participation	0	0.00	5	11.97	1	0.99	6	2.54
Confidentiality or survey legitimacy concerns	0	0.00	2	3.76	3	5.89	5	5.21
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	1	0.95	0	0.00	1	0.14

Table 7.24 2012 Interview Refusal Reasons, by Age (Arkansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	70	100.00	79	100.00	105	100.00	254	100.00
Parental refusal	48	68.57	0	0.00	0	0.00	48	18.90
Nothing in it for me	14	20.00	33	41.77	57	54.29	104	40.94
No time	5	7.14	36	45.57	37	35.24	78	30.71
Government/surveys too invasive	1	1.43	4	5.06	0	0.00	5	1.97
Gatekeeper/household member won't allow participation	2	2.86	6	7.59	5	4.76	13	5.12
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	4	3.81	4	1.57
House too messy/too ill	0	0.00	0	0.00	2	1.90	2	0.79
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Arkansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	70	100.00	79	100.00	105	100.00	254	100.00
Parental refusal	48	66.08	0	0.00	0	0.00	48	4.76
Nothing in it for me	14	21.28	33	38.53	57	55.57	104	51.16
No time	5	8.38	36	49.85	37	32.19	78	32.49
Government/surveys too invasive	1	1.42	4	5.26	0	0.00	5	0.70
Gatekeeper/household member won't allow participation	2	2.84	6	6.36	5	4.19	13	4.34
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	4	5.03	4	4.09
House too messy/too ill	0	0.00	0	0.00	2	3.02	2	2.45
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (California) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	215	100.00	296	100.00	425	100.00	936	100.00
Parental refusal	145	67.44	0	0.00	0	0.00	145	15.49
Nothing in it for me	47	21.86	178	60.14	263	61.88	488	52.14
No time	9	4.19	48	16.22	76	17.88	133	14.21
Government/surveys too invasive	7	3.26	11	3.72	30	7.06	48	5.13
Gatekeeper/household member won't allow participation	5	2.33	38	12.84	19	4.47	62	6.62
Confidentiality or survey legitimacy concerns	2	0.93	10	3.38	30	7.06	42	4.49
House too messy/too ill	0	0.00	2	0.68	3	0.71	5	0.53
Other	0	0.00	9	3.04	4	0.94	13	1.39
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (California) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	215	100.00	296	100.00	425	100.00	936	100.00
Parental refusal	145	63.41	0	0.00	0	0.00	145	4.32
Nothing in it for me	47	22.46	178	58.70	263	60.81	488	57.96
No time	9	4.05	48	16.35	76	17.51	133	16.46
Government/surveys too invasive	7	2.71	11	3.61	30	7.87	48	7.03
Gatekeeper/household member won't allow participation	5	3.03	38	13.18	19	4.17	62	5.13
Confidentiality or survey legitimacy concerns	2	4.33	10	2.60	30	7.09	42	6.39
House too messy/too ill	0	0.00	2	1.22	3	0.93	5	0.90
Other	0	0.00	9	4.33	4	1.61	13	1.81
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Colorado) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	46	100.00	58	100.00	86	100.00	190	100.00
Parental refusal	36	78.26	0	0.00	0	0.00	36	18.95
Nothing in it for me	7	15.22	35	60.34	48	55.81	90	47.37
No time	1	2.17	9	15.52	20	23.26	30	15.79
Government/surveys too invasive	0	0.00	4	6.90	16	18.60	20	10.53
Gatekeeper/household member won't allow participation	2	4.35	8	13.79	1	1.16	11	5.79
Confidentiality or survey legitimacy concerns	0	0.00	2	3.45	1	1.16	3	1.58
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Colorado) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	46	100.00	58	100.00	86	100.00	190	100.00
Parental refusal	36	75.26	0	0.00	0	0.00	36	3.98
Nothing in it for me	7	15.19	35	57.53	48	57.83	90	55.54
No time	1	3.16	9	12.88	20	17.66	30	16.39
Government/surveys too invasive	0	0.00	4	8.14	16	21.71	20	19.14
Gatekeeper/household member won't allow participation	2	6.38	8	17.35	1	1.93	11	3.78
Confidentiality or survey legitimacy concerns	0	0.00	2	4.09	1	0.88	3	1.17
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Connecticut) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	64	100.00	65	100.00	98	100.00	227	100.00
Parental refusal	51	79.69	0	0.00	0	0.00	51	22.47
Nothing in it for me	8	12.50	36	55.38	49	50.00	93	40.97
No time	4	6.25	12	18.46	25	25.51	41	18.06
Government/surveys too invasive	0	0.00	4	6.15	9	9.18	13	5.73
Gatekeeper/household member won't allow participation	1	1.56	12	18.46	12	12.24	25	11.01
Confidentiality or survey legitimacy concerns	0	0.00	1	1.54	3	3.06	4	1.76
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Connecticut) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	64	100.00	65	100.00	98	100.00	227	100.00
Parental refusal	51	78.80	0	0.00	0	0.00	51	6.02
Nothing in it for me	8	14.14	36	52.95	49	50.04	93	47.53
No time	4	5.87	12	20.62	25	27.07	41	24.94
Government/surveys too invasive	0	0.00	4	5.83	9	10.38	13	9.22
Gatekeeper/household member won't allow participation	1	1.19	12	18.69	12	10.24	25	10.22
Confidentiality or survey legitimacy concerns	0	0.00	1	1.91	3	2.27	4	2.07
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Delaware) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	65	100.00	48	100.00	80	100.00	193	100.00
Parental refusal	55	84.62	0	0.00	0	0.00	55	28.50
Nothing in it for me	3	4.62	12	25.00	26	32.50	41	21.24
No time	1	1.54	13	27.08	26	32.50	40	20.73
Government/surveys too invasive	3	4.62	3	6.25	13	16.25	19	9.84
Gatekeeper/household member won't allow participation	0	0.00	14	29.17	6	7.50	20	10.36
Confidentiality or survey legitimacy concerns	3	4.62	6	12.50	9	11.25	18	9.33
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Delaware) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	65	100.00	48	100.00	80	100.00	193	100.00
Parental refusal	55	85.41	0	0.00	0	0.00	55	6.37
Nothing in it for me	3	3.73	12	23.29	26	34.14	41	30.92
No time	1	1.99	13	23.00	26	29.86	40	27.18
Government/surveys too invasive	3	4.47	3	8.08	13	17.72	19	15.88
Gatekeeper/household member won't allow participation	0	0.00	14	29.05	6	8.13	20	9.37
Confidentiality or survey legitimacy concerns	3	4.40	6	16.58	9	10.15	18	10.29
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (District of Columbia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	25	100.00	40	100.00	53	100.00	118	100.00
Parental refusal	18	72.00	0	0.00	0	0.00	18	15.25
Nothing in it for me	3	12.00	13	32.50	24	45.28	40	33.90
No time	1	4.00	16	40.00	21	39.62	38	32.20
Government/surveys too invasive	1	4.00	2	5.00	6	11.32	9	7.63
Gatekeeper/household member won't allow participation	2	8.00	6	15.00	2	3.77	10	8.47
Confidentiality or survey legitimacy concerns	0	0.00	3	7.50	0	0.00	3	2.54
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (District of Columbia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	25	100.00	40	100.00	53	100.00	118	100.00
Parental refusal	18	69.69	0	0.00	0	0.00	18	1.83
Nothing in it for me	3	13.25	13	34.44	24	43.20	40	41.36
No time	1	5.54	16	35.65	21	42.21	38	40.46
Government/surveys too invasive	1	4.35	2	5.00	6	11.75	9	10.74
Gatekeeper/household member won't allow participation	2	7.17	6	17.99	2	2.84	10	4.77
Confidentiality or survey legitimacy concerns	0	0.00	3	6.92	0	0.00	3	0.83
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Florida) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	160	100.00	217	100.00	359	100.00	736	100.00
Parental refusal	131	81.88	0	0.00	0	0.00	131	17.80
Nothing in it for me	15	9.38	95	43.78	153	42.62	263	35.73
No time	6	3.75	65	29.95	110	30.64	181	24.59
Government/surveys too invasive	3	1.88	12	5.53	33	9.19	48	6.52
Gatekeeper/household member won't allow participation	4	2.50	21	9.68	9	2.51	34	4.62
Confidentiality or survey legitimacy concerns	0	0.00	19	8.76	38	10.58	57	7.74
House too messy/too ill	1	0.63	1	0.46	7	1.95	9	1.22
Other	0	0.00	4	1.84	8	2.23	12	1.63
Missing	0	0.00	0	0.00	1	0.28	1	0.14

Table 7.25 2012 Interview Refusal Reasons, by Age (Florida) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	160	100.00	217	100.00	359	100.00	736	100.00
Parental refusal	131	83.51	0	0.00	0	0.00	131	3.96
Nothing in it for me	15	7.64	95	42.52	153	41.87	263	40.30
No time	6	4.06	65	30.85	110	28.92	181	27.90
Government/surveys too invasive	3	2.26	12	5.77	33	10.71	48	9.90
Gatekeeper/household member won't allow participation	4	2.00	21	9.94	9	2.48	34	3.07
Confidentiality or survey legitimacy concerns	0	0.00	19	8.62	38	11.20	57	10.46
House too messy/too ill	1	0.54	1	0.38	7	2.11	9	1.89
Other	0	0.00	4	1.92	8	2.52	12	2.35
Missing	0	0.00	0	0.00	1	0.20	1	0.17

Table 7.24 2012 Interview Refusal Reasons, by Age (Georgia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	39	100.00	53	100.00	86	100.00	178	100.00
Parental refusal	17	43.59	0	0.00	0	0.00	17	9.55
Nothing in it for me	6	15.38	19	35.85	25	29.07	50	28.09
No time	4	10.26	18	33.96	33	38.37	55	30.90
Government/surveys too invasive	6	15.38	4	7.55	15	17.44	25	14.04
Gatekeeper/household member won't allow participation	0	0.00	4	7.55	1	1.16	5	2.81
Confidentiality or survey legitimacy concerns	4	10.26	4	7.55	11	12.79	19	10.67
House too messy/too ill	0	0.00	1	1.89	0	0.00	1	0.56
Other	2	5.13	3	5.66	1	1.16	6	3.37
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Georgia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	39	100.00	53	100.00	86	100.00	178	100.00
Parental refusal	17	52.71	0	0.00	0	0.00	17	3.60
Nothing in it for me	6	12.83	19	33.82	25	32.53	50	31.32
No time	4	7.09	18	32.09	33	30.26	55	28.86
Government/surveys too invasive	6	12.78	4	6.99	15	19.07	25	17.44
Gatekeeper/household member won't allow participation	0	0.00	4	6.94	1	1.49	5	1.93
Confidentiality or survey legitimacy concerns	4	10.83	4	14.44	11	16.34	19	15.77
House too messy/too ill	0	0.00	1	1.83	0	0.00	1	0.18
Other	2	3.75	3	3.89	1	0.32	6	0.91
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Hawaii) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	72	100.00	52	100.00	124	100.00	248	100.00
Parental refusal	61	84.72	0	0.00	0	0.00	61	24.60
Nothing in it for me	6	8.33	20	38.46	63	50.81	89	35.89
No time	1	1.39	18	34.62	37	29.84	56	22.58
Government/surveys too invasive	0	0.00	2	3.85	8	6.45	10	4.03
Gatekeeper/household member won't allow participation	3	4.17	6	11.54	7	5.65	16	6.45
Confidentiality or survey legitimacy concerns	0	0.00	3	5.77	3	2.42	6	2.42
House too messy/too ill	0	0.00	0	0.00	2	1.61	2	0.81
Other	1	1.39	2	3.85	3	2.42	6	2.42
Missing	0	0.00	1	1.92	1	0.81	2	0.81

Table 7.25 2012 Interview Refusal Reasons, by Age (Hawaii) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	72	100.00	52	100.00	124	100.00	248	100.00
Parental refusal	61	83.13	0	0.00	0	0.00	61	6.06
Nothing in it for me	6	10.52	20	42.54	63	52.54	89	48.73
No time	1	1.63	18	31.05	37	27.28	56	25.69
Government/surveys too invasive	0	0.00	2	3.02	8	7.05	10	6.24
Gatekeeper/household member won't allow participation	3	3.33	6	11.44	7	6.49	16	6.63
Confidentiality or survey legitimacy concerns	0	0.00	3	6.58	3	2.53	6	2.65
House too messy/too ill	0	0.00	0	0.00	2	1.45	2	1.24
Other	1	1.40	2	3.89	3	1.76	6	1.89
Missing	0	0.00	1	1.48	1	0.89	2	0.87

Table 7.24 2012 Interview Refusal Reasons, by Age (Idaho) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	63	100.00	89	100.00	187	100.00
Parental refusal	20	57.14	0	0.00	0	0.00	20	10.70
Nothing in it for me	6	17.14	33	52.38	41	46.07	80	42.78
No time	7	20.00	17	26.98	24	26.97	48	25.67
Government/surveys too invasive	2	5.71	4	6.35	17	19.10	23	12.30
Gatekeeper/household member won't allow participation	0	0.00	9	14.29	1	1.12	10	5.35
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	6	6.74	6	3.21
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Idaho) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	63	100.00	89	100.00	187	100.00
Parental refusal	20	56.91	0	0.00	0	0.00	20	3.37
Nothing in it for me	6	15.72	33	51.57	41	45.15	80	44.30
No time	7	21.05	17	20.03	24	24.56	48	23.72
Government/surveys too invasive	2	6.32	4	11.18	17	19.79	23	17.80
Gatekeeper/household member won't allow participation	0	0.00	9	17.22	1	1.33	10	3.46
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	6	9.17	6	7.35
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Illinois) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	225	100.00	240	100.00	379	100.00	844	100.00
Parental refusal	154	68.44	0	0.00	0	0.00	154	18.25
Nothing in it for me	46	20.44	133	55.42	219	57.78	398	47.16
No time	11	4.89	59	24.58	99	26.12	169	20.02
Government/surveys too invasive	2	0.89	6	2.50	22	5.80	30	3.55
Gatekeeper/household member won't allow participation	11	4.89	29	12.08	14	3.69	54	6.40
Confidentiality or survey legitimacy concerns	1	0.44	6	2.50	20	5.28	27	3.20
House too messy/too ill	0	0.00	0	0.00	2	0.53	2	0.24
Other	0	0.00	5	2.08	2	0.53	7	0.83
Missing	0	0.00	2	0.83	1	0.26	3	0.36

Table 7.25 2012 Interview Refusal Reasons, by Age (Illinois) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	225	100.00	240	100.00	379	100.00	844	100.00
Parental refusal	154	68.47	0	0.00	0	0.00	154	4.89
Nothing in it for me	46	20.13	133	55.08	219	58.80	398	55.68
No time	11	4.30	59	23.93	99	23.98	169	22.57
Government/surveys too invasive	2	0.95	6	2.42	22	7.13	30	6.24
Gatekeeper/household member won't allow participation	11	5.73	29	13.04	14	3.59	54	4.65
Confidentiality or survey legitimacy concerns	1	0.41	6	2.47	20	4.96	27	4.40
House too messy/too ill	0	0.00	0	0.00	2	0.56	2	0.46
Other	0	0.00	5	2.06	2	0.71	7	0.79
Missing	0	0.00	2	1.01	1	0.27	3	0.32

Table 7.24 2012 Interview Refusal Reasons, by Age (Indiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	48	100.00	60	100.00	99	100.00	207	100.00
Parental refusal	43	89.58	0	0.00	0	0.00	43	20.77
Nothing in it for me	3	6.25	23	38.33	35	35.35	61	29.47
No time	0	0.00	15	25.00	26	26.26	41	19.81
Government/surveys too invasive	2	4.17	11	18.33	28	28.28	41	19.81
Gatekeeper/household member won't allow participation	0	0.00	11	18.33	6	6.06	17	8.21
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	3	3.03	3	1.45
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	1.01	1	0.48
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Indiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	48	100.00	60	100.00	99	100.00	207	100.00
Parental refusal	43	89.14	0	0.00	0	0.00	43	5.27
Nothing in it for me	3	7.11	23	37.01	35	34.31	61	32.93
No time	0	0.00	15	23.35	26	29.07	41	26.88
Government/surveys too invasive	2	3.75	11	18.63	28	27.57	41	25.43
Gatekeeper/household member won't allow participation	0	0.00	11	21.02	6	5.74	17	6.67
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	3	2.44	3	2.09
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	0.87	1	0.74
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Iowa) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	50	100.00	42	100.00	85	100.00	177	100.00
Parental refusal	42	84.00	0	0.00	0	0.00	42	23.73
Nothing in it for me	4	8.00	23	54.76	50	58.82	77	43.50
No time	1	2.00	11	26.19	11	12.94	23	12.99
Government/surveys too invasive	1	2.00	2	4.76	10	11.76	13	7.34
Gatekeeper/household member won't allow participation	2	4.00	6	14.29	10	11.76	18	10.17
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.18	1	0.56
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	2	2.35	2	1.13
Missing	0	0.00	0	0.00	1	1.18	1	0.56

Table 7.25 2012 Interview Refusal Reasons, by Age (Iowa) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	50	100.00	42	100.00	85	100.00	177	100.00
Parental refusal	42	84.44	0	0.00	0	0.00	42	5.62
Nothing in it for me	4	7.99	23	49.52	50	56.48	77	52.68
No time	1	1.50	11	33.01	11	14.09	23	14.81
Government/surveys too invasive	1	1.39	2	4.29	10	14.22	13	12.55
Gatekeeper/household member won't allow participation	2	4.68	6	13.17	10	11.16	18	10.90
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	0.82	1	0.70
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	2	1.93	2	1.64
Missing	0	0.00	0	0.00	1	1.30	1	1.11

Table 7.24 2012 Interview Refusal Reasons, by Age (Kansas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	34	100.00	77	100.00	148	100.00
Parental refusal	24	64.86	0	0.00	0	0.00	24	16.22
Nothing in it for me	9	24.32	14	41.18	44	57.14	67	45.27
No time	1	2.70	13	38.24	26	33.77	40	27.03
Government/surveys too invasive	0	0.00	1	2.94	1	1.30	2	1.35
Gatekeeper/household member won't allow participation	3	8.11	5	14.71	5	6.49	13	8.78
Confidentiality or survey legitimacy concerns	0	0.00	1	2.94	0	0.00	1	0.68
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	1.30	1	0.68
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Kansas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	34	100.00	77	100.00	148	100.00
Parental refusal	24	67.68	0	0.00	0	0.00	24	4.12
Nothing in it for me	9	21.30	14	39.90	44	60.27	67	56.38
No time	1	2.90	13	40.25	26	29.95	40	29.07
Government/surveys too invasive	0	0.00	1	2.69	1	0.80	2	0.89
Gatekeeper/household member won't allow participation	3	8.12	5	13.87	5	7.71	13	8.19
Confidentiality or survey legitimacy concerns	0	0.00	1	3.28	0	0.00	1	0.25
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	1.28	1	1.10
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Kentucky) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	49	100.00	43	100.00	82	100.00	174	100.00
Parental refusal	36	73.47	0	0.00	0	0.00	36	20.69
Nothing in it for me	6	12.24	8	18.60	37	45.12	51	29.31
No time	4	8.16	26	60.47	28	34.15	58	33.33
Government/surveys too invasive	3	6.12	2	4.65	7	8.54	12	6.90
Gatekeeper/household member won't allow participation	0	0.00	6	13.95	1	1.22	7	4.02
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	5	6.10	5	2.87
House too messy/too ill	0	0.00	0	0.00	3	3.66	3	1.72
Other	0	0.00	1	2.33	1	1.22	2	1.15
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Kentucky) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	49	100.00	43	100.00	82	100.00	174	100.00
Parental refusal	36	73.09	0	0.00	0	0.00	36	4.91
Nothing in it for me	6	11.84	8	17.17	37	46.17	51	41.95
No time	4	8.42	26	63.92	28	28.00	58	29.06
Government/surveys too invasive	3	6.65	2	4.30	7	10.70	12	10.01
Gatekeeper/household member won't allow participation	0	0.00	6	12.37	1	1.03	7	1.71
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	5	8.54	5	7.40
House too messy/too ill	0	0.00	0	0.00	3	4.90	3	4.25
Other	0	0.00	1	2.25	1	0.66	2	0.72
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Louisiana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	29	100.00	36	100.00	76	100.00	141	100.00
Parental refusal	24	82.76	0	0.00	0	0.00	24	17.02
Nothing in it for me	3	10.34	19	52.78	47	61.84	69	48.94
No time	0	0.00	10	27.78	20	26.32	30	21.28
Government/surveys too invasive	1	3.45	2	5.56	5	6.58	8	5.67
Gatekeeper/household member won't allow participation	1	3.45	5	13.89	1	1.32	7	4.96
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.32	1	0.71
House too messy/too ill	0	0.00	0	0.00	1	1.32	1	0.71
Other	0	0.00	0	0.00	1	1.32	1	0.71
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Louisiana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	29	100.00	36	100.00	76	100.00	141	100.00
Parental refusal	24	82.33	0	0.00	0	0.00	24	4.33
Nothing in it for me	3	11.09	19	47.10	47	61.59	69	57.71
No time	0	0.00	10	31.22	20	23.79	30	23.17
Government/surveys too invasive	1	3.17	2	5.42	5	9.34	8	8.69
Gatekeeper/household member won't allow participation	1	3.42	5	16.26	1	0.94	7	2.36
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	0.92	1	0.79
House too messy/too ill	0	0.00	0	0.00	1	1.76	1	1.52
Other	0	0.00	0	0.00	1	1.65	1	1.43
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Maine) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	44	100.00	44	100.00	63	100.00	151	100.00
Parental refusal	34	77.27	0	0.00	0	0.00	34	22.52
Nothing in it for me	6	13.64	29	65.91	31	49.21	66	43.71
No time	3	6.82	7	15.91	12	19.05	22	14.57
Government/surveys too invasive	0	0.00	3	6.82	10	15.87	13	8.61
Gatekeeper/household member won't allow participation	1	2.27	4	9.09	5	7.94	10	6.62
Confidentiality or survey legitimacy concerns	0	0.00	1	2.27	4	6.35	5	3.31
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	1.59	1	0.66
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Maine) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	44	100.00	44	100.00	63	100.00	151	100.00
Parental refusal	34	76.64	0	0.00	0	0.00	34	5.17
Nothing in it for me	6	13.32	29	69.69	31	45.23	66	44.93
No time	3	7.78	7	13.88	12	19.79	22	18.53
Government/surveys too invasive	0	0.00	3	4.50	10	18.71	13	16.37
Gatekeeper/household member won't allow participation	1	2.25	4	9.33	5	7.74	10	7.49
Confidentiality or survey legitimacy concerns	0	0.00	1	2.59	4	7.58	5	6.69
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	1	0.95	1	0.82
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Maryland) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	43	100.00	44	100.00	76	100.00	163	100.00
Parental refusal	35	81.40	0	0.00	0	0.00	35	21.47
Nothing in it for me	4	9.30	19	43.18	42	55.26	65	39.88
No time	2	4.65	9	20.45	16	21.05	27	16.56
Government/surveys too invasive	2	4.65	3	6.82	13	17.11	18	11.04
Gatekeeper/household member won't allow participation	0	0.00	13	29.55	5	6.58	18	11.04
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Maryland) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	43	100.00	44	100.00	76	100.00	163	100.00
Parental refusal	35	83.88	0	0.00	0	0.00	35	5.71
Nothing in it for me	4	8.99	19	44.26	42	55.24	65	51.07
No time	2	4.37	9	22.27	16	19.23	27	18.50
Government/surveys too invasive	2	2.77	3	6.11	13	18.05	18	15.90
Gatekeeper/household member won't allow participation	0	0.00	13	27.36	5	7.48	18	8.82
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Massachusetts) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	60	100.00	49	100.00	83	100.00	192	100.00
Parental refusal	45	75.00	0	0.00	0	0.00	45	23.44
Nothing in it for me	7	11.67	20	40.82	27	32.53	54	28.13
No time	5	8.33	12	24.49	33	39.76	50	26.04
Government/surveys too invasive	1	1.67	4	8.16	9	10.84	14	7.29
Gatekeeper/household member won't allow participation	0	0.00	11	22.45	8	9.64	19	9.90
Confidentiality or survey legitimacy concerns	2	3.33	1	2.04	2	2.41	5	2.60
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	1	2.04	3	3.61	4	2.08
Missing	0	0.00	0	0.00	1	1.20	1	0.52

Table 7.25 2012 Interview Refusal Reasons, by Age (Massachusetts) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	60	100.00	49	100.00	83	100.00	192	100.00
Parental refusal	45	69.03	0	0.00	0	0.00	45	5.42
Nothing in it for me	7	11.89	20	43.00	27	36.42	54	35.11
No time	5	7.88	12	23.30	33	37.70	50	34.01
Government/surveys too invasive	1	1.56	4	6.56	9	10.34	14	9.29
Gatekeeper/household member won't allow participation	0	0.00	11	23.53	8	7.79	19	8.65
Confidentiality or survey legitimacy concerns	2	9.65	1	1.83	2	3.32	5	3.68
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	1	1.77	3	3.75	4	3.27
Missing	0	0.00	0	0.00	1	0.69	1	0.57

Table 7.24 2012 Interview Refusal Reasons, by Age (Michigan) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	224	100.00	196	100.00	334	100.00	754	100.00
Parental refusal	178	79.46	0	0.00	0	0.00	178	23.61
Nothing in it for me	22	9.82	104	53.06	187	55.99	313	41.51
No time	12	5.36	43	21.94	93	27.84	148	19.63
Government/surveys too invasive	4	1.79	10	5.10	21	6.29	35	4.64
Gatekeeper/household member won't allow participation	7	3.13	34	17.35	11	3.29	52	6.90
Confidentiality or survey legitimacy concerns	1	0.45	2	1.02	12	3.59	15	1.99
House too messy/too ill	0	0.00	0	0.00	3	0.90	3	0.40
Other	0	0.00	3	1.53	5	1.50	8	1.06
Missing	0	0.00	0	0.00	2	0.60	2	0.27

Table 7.25 2012 Interview Refusal Reasons, by Age (Michigan) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	224	100.00	196	100.00	334	100.00	754	100.00
Parental refusal	178	80.15	0	0.00	0	0.00	178	6.17
Nothing in it for me	22	8.52	104	53.73	187	55.77	313	51.95
No time	12	5.32	43	20.92	93	26.89	148	24.69
Government/surveys too invasive	4	1.65	10	5.30	21	6.57	35	6.08
Gatekeeper/household member won't allow participation	7	3.28	34	17.80	11	3.26	52	4.57
Confidentiality or survey legitimacy concerns	1	1.08	2	0.91	12	3.93	15	3.44
House too messy/too ill	0	0.00	0	0.00	3	1.12	3	0.94
Other	0	0.00	3	1.34	5	1.56	8	1.42
Missing	0	0.00	0	0.00	2	0.89	2	0.74

Table 7.24 2012 Interview Refusal Reasons, by Age (Minnesota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	48	100.00	62	100.00	145	100.00
Parental refusal	30	85.71	0	0.00	0	0.00	30	20.69
Nothing in it for me	2	5.71	15	31.25	31	50.00	48	33.10
No time	1	2.86	16	33.33	17	27.42	34	23.45
Government/surveys too invasive	1	2.86	4	8.33	6	9.68	11	7.59
Gatekeeper/household member won't allow participation	1	2.86	7	14.58	4	6.45	12	8.28
Confidentiality or survey legitimacy concerns	0	0.00	6	12.50	3	4.84	9	6.21
House too messy/too ill	0	0.00	0	0.00	1	1.61	1	0.69
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Minnesota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	48	100.00	62	100.00	145	100.00
Parental refusal	30	86.32	0	0.00	0	0.00	30	5.45
Nothing in it for me	2	4.42	15	34.27	31	48.59	48	43.76
No time	1	2.71	16	34.52	17	28.14	34	27.44
Government/surveys too invasive	1	3.93	4	7.48	6	5.83	11	5.94
Gatekeeper/household member won't allow participation	1	2.62	7	13.50	4	8.13	12	8.55
Confidentiality or survey legitimacy concerns	0	0.00	6	10.23	3	6.66	9	6.75
House too messy/too ill	0	0.00	0	0.00	1	2.65	1	2.11
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Mississippi) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	48	100.00	15	100.00	52	100.00	115	100.00
Parental refusal	37	77.08	0	0.00	0	0.00	37	32.17
Nothing in it for me	8	16.67	5	33.33	28	53.85	41	35.65
No time	0	0.00	3	20.00	13	25.00	16	13.91
Government/surveys too invasive	2	4.17	2	13.33	6	11.54	10	8.70
Gatekeeper/household member won't allow participation	1	2.08	4	26.67	2	3.85	7	6.09
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	3.85	2	1.74
House too messy/too ill	0	0.00	0	0.00	1	1.92	1	0.87
Other	0	0.00	1	6.67	0	0.00	1	0.87
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Mississippi) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	48	100.00	15	100.00	52	100.00	115	100.00
Parental refusal	37	70.71	0	0.00	0	0.00	37	7.70
Nothing in it for me	8	17.37	5	39.44	28	50.31	41	46.19
No time	0	0.00	3	21.45	13	26.10	16	23.03
Government/surveys too invasive	2	10.29	2	10.63	6	14.38	10	13.75
Gatekeeper/household member won't allow participation	1	1.62	4	25.16	2	3.15	7	4.06
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	3.08	2	2.60
House too messy/too ill	0	0.00	0	0.00	1	2.98	1	2.51
Other	0	0.00	1	3.32	0	0.00	1	0.16
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Missouri) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	48	100.00	92	100.00	177	100.00
Parental refusal	32	86.49	0	0.00	0	0.00	32	18.08
Nothing in it for me	2	5.41	25	52.08	49	53.26	76	42.94
No time	0	0.00	9	18.75	13	14.13	22	12.43
Government/surveys too invasive	2	5.41	5	10.42	12	13.04	19	10.73
Gatekeeper/household member won't allow participation	0	0.00	9	18.75	6	6.52	15	8.47
Confidentiality or survey legitimacy concerns	1	2.70	0	0.00	8	8.70	9	5.08
House too messy/too ill	0	0.00	0	0.00	1	1.09	1	0.56
Other	0	0.00	0	0.00	3	3.26	3	1.69
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Missouri) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	48	100.00	92	100.00	177	100.00
Parental refusal	32	84.64	0	0.00	0	0.00	32	3.76
Nothing in it for me	2	5.04	25	51.79	49	51.35	76	49.33
No time	0	0.00	9	18.65	13	14.04	22	13.79
Government/surveys too invasive	2	7.63	5	9.68	12	14.75	19	14.02
Gatekeeper/household member won't allow participation	0	0.00	9	19.89	6	5.62	15	6.54
Confidentiality or survey legitimacy concerns	1	2.69	0	0.00	8	9.65	9	8.55
House too messy/too ill	0	0.00	0	0.00	1	1.35	1	1.18
Other	0	0.00	0	0.00	3	3.25	3	2.84
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Montana) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	67	100.00	59	100.00	81	100.00	207	100.00
Parental refusal	56	83.58	0	0.00	0	0.00	56	27.05
Nothing in it for me	4	5.97	31	52.54	45	55.56	80	38.65
No time	5	7.46	20	33.90	22	27.16	47	22.71
Government/surveys too invasive	2	2.99	2	3.39	13	16.05	17	8.21
Gatekeeper/household member won't allow participation	0	0.00	6	10.17	0	0.00	6	2.90
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	1	1.23	1	0.48
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Montana) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	67	100.00	59	100.00	81	100.00	207	100.00
Parental refusal	56	84.46	0	0.00	0	0.00	56	7.00
Nothing in it for me	4	4.44	31	52.72	45	50.48	80	46.95
No time	5	6.78	20	32.20	22	28.48	47	27.15
Government/surveys too invasive	2	4.32	2	3.16	13	19.22	17	15.98
Gatekeeper/household member won't allow participation	0	0.00	6	11.92	0	0.00	6	1.49
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	1	1.81	1	1.44
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Nebraska) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	34	100.00	96	100.00	165	100.00
Parental refusal	30	85.71	0	0.00	0	0.00	30	18.18
Nothing in it for me	2	5.71	11	32.35	34	35.42	47	28.48
No time	1	2.86	13	38.24	33	34.38	47	28.48
Government/surveys too invasive	1	2.86	6	17.65	23	23.96	30	18.18
Gatekeeper/household member won't allow participation	0	0.00	4	11.76	0	0.00	4	2.42
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	2.08	2	1.21
House too messy/too ill	0	0.00	0	0.00	2	2.08	2	1.21
Other	1	2.86	0	0.00	1	1.04	2	1.21
Missing	0	0.00	0	0.00	1	1.04	1	0.61

Table 7.25 2012 Interview Refusal Reasons, by Age (Nebraska) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	34	100.00	96	100.00	165	100.00
Parental refusal	30	87.53	0	0.00	0	0.00	30	4.39
Nothing in it for me	2	4.51	11	32.37	34	38.30	47	36.35
No time	1	2.79	13	38.38	33	27.44	47	26.68
Government/surveys too invasive	1	2.83	6	18.82	23	27.47	30	25.86
Gatekeeper/household member won't allow participation	0	0.00	4	10.43	0	0.00	4	0.45
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	1.88	2	1.70
House too messy/too ill	0	0.00	0	0.00	2	2.45	2	2.22
Other	1	2.34	0	0.00	1	1.25	2	1.25
Missing	0	0.00	0	0.00	1	1.21	1	1.10

Table 7.24 2012 Interview Refusal Reasons, by Age (Nevada) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	63	100.00	89	100.00	189	100.00
Parental refusal	22	59.46	0	0.00	0	0.00	22	11.64
Nothing in it for me	5	13.51	32	50.79	36	40.45	73	38.62
No time	3	8.11	17	26.98	30	33.71	50	26.46
Government/surveys too invasive	0	0.00	5	7.94	13	14.61	18	9.52
Gatekeeper/household member won't allow participation	4	10.81	4	6.35	2	2.25	10	5.29
Confidentiality or survey legitimacy concerns	0	0.00	3	4.76	2	2.25	5	2.65
House too messy/too ill	0	0.00	0	0.00	1	1.12	1	0.53
Other	3	8.11	2	3.17	5	5.62	10	5.29
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Nevada) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	63	100.00	89	100.00	189	100.00
Parental refusal	22	52.94	0	0.00	0	0.00	22	2.81
Nothing in it for me	5	11.23	32	44.65	36	37.50	73	36.80
No time	3	12.55	17	30.18	30	41.27	50	38.67
Government/surveys too invasive	0	0.00	5	10.42	13	12.36	18	11.52
Gatekeeper/household member won't allow participation	4	18.37	4	5.93	2	2.11	10	3.34
Confidentiality or survey legitimacy concerns	0	0.00	3	5.83	2	1.44	5	1.79
House too messy/too ill	0	0.00	0	0.00	1	0.90	1	0.77
Other	3	4.91	2	2.98	5	4.41	10	4.30
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (New Hampshire) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	87	100.00	75	100.00	102	100.00	264	100.00
Parental refusal	77	88.51	0	0.00	0	0.00	77	29.17
Nothing in it for me	10	11.49	46	61.33	55	53.92	111	42.05
No time	0	0.00	13	17.33	35	34.31	48	18.18
Government/surveys too invasive	0	0.00	3	4.00	6	5.88	9	3.41
Gatekeeper/household member won't allow participation	0	0.00	10	13.33	4	3.92	14	5.30
Confidentiality or survey legitimacy concerns	0	0.00	1	1.33	2	1.96	3	1.14
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	2	2.67	0	0.00	2	0.76
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (New Hampshire) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	87	100.00	75	100.00	102	100.00	264	100.00
Parental refusal	77	88.43	0	0.00	0	0.00	77	7.62
Nothing in it for me	10	11.57	46	64.56	55	53.42	111	50.85
No time	0	0.00	13	16.42	35	32.56	48	28.25
Government/surveys too invasive	0	0.00	3	3.68	6	7.34	9	6.36
Gatekeeper/household member won't allow participation	0	0.00	10	11.99	4	5.35	14	5.51
Confidentiality or survey legitimacy concerns	0	0.00	1	2.80	2	1.33	3	1.36
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	2	0.54	0	0.00	2	0.05
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (New Jersey) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	55	100.00	69	100.00	92	100.00	216	100.00
Parental refusal	48	87.27	0	0.00	0	0.00	48	22.22
Nothing in it for me	2	3.64	36	52.17	48	52.17	86	39.81
No time	3	5.45	18	26.09	29	31.52	50	23.15
Government/surveys too invasive	0	0.00	3	4.35	3	3.26	6	2.78
Gatekeeper/household member won't allow participation	2	3.64	11	15.94	6	6.52	19	8.80
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	3	3.26	3	1.39
House too messy/too ill	0	0.00	0	0.00	3	3.26	3	1.39
Other	0	0.00	1	1.45	0	0.00	1	0.46
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (New Jersey) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	55	100.00	69	100.00	92	100.00	216	100.00
Parental refusal	48	88.86	0	0.00	0	0.00	48	6.31
Nothing in it for me	2	2.38	36	55.92	48	54.33	86	50.81
No time	3	5.98	18	25.35	29	28.80	50	26.83
Government/surveys too invasive	0	0.00	3	2.83	3	2.04	6	1.98
Gatekeeper/household member won't allow participation	2	2.77	11	15.18	6	7.81	19	8.20
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	3	4.18	3	3.46
House too messy/too ill	0	0.00	0	0.00	3	2.84	3	2.35
Other	0	0.00	1	0.72	0	0.00	1	0.07
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (New Mexico) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	53	100.00	94	100.00	184	100.00
Parental refusal	25	67.57	0	0.00	0	0.00	25	13.59
Nothing in it for me	7	18.92	28	52.83	46	48.94	81	44.02
No time	1	2.70	16	30.19	38	40.43	55	29.89
Government/surveys too invasive	2	5.41	5	9.43	6	6.38	13	7.07
Gatekeeper/household member won't allow participation	2	5.41	2	3.77	0	0.00	4	2.17
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.06	1	0.54
House too messy/too ill	0	0.00	0	0.00	1	1.06	1	0.54
Other	0	0.00	2	3.77	2	2.13	4	2.17
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (New Mexico) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	53	100.00	94	100.00	184	100.00
Parental refusal	25	70.17	0	0.00	0	0.00	25	4.17
Nothing in it for me	7	15.17	28	50.65	46	46.09	81	44.71
No time	1	2.32	16	29.66	38	41.72	55	38.16
Government/surveys too invasive	2	5.14	5	10.14	6	7.17	13	7.35
Gatekeeper/household member won't allow participation	2	7.20	2	4.30	0	0.00	4	0.86
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	0.65	1	0.54
House too messy/too ill	0	0.00	0	0.00	1	1.70	1	1.42
Other	0	0.00	2	5.26	2	2.69	4	2.79
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (New York) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	292	100.00	319	100.00	470	100.00	1,081	100.00
Parental refusal	242	82.88	0	0.00	0	0.00	242	22.39
Nothing in it for me	33	11.30	165	51.72	226	48.09	424	39.22
No time	7	2.40	64	20.06	137	29.15	208	19.24
Government/surveys too invasive	4	1.37	18	5.64	40	8.51	62	5.74
Gatekeeper/household member won't allow participation	5	1.71	58	18.18	32	6.81	95	8.79
Confidentiality or survey legitimacy concerns	1	0.34	10	3.13	26	5.53	37	3.42
House too messy/too ill	0	0.00	0	0.00	6	1.28	6	0.56
Other	0	0.00	4	1.25	3	0.64	7	0.65
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (New York) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	292	100.00	319	100.00	470	100.00	1,081	100.00
Parental refusal	242	82.33	0	0.00	0	0.00	242	5.95
Nothing in it for me	33	11.31	165	51.08	226	50.26	424	47.53
No time	7	2.42	64	18.90	137	27.89	208	25.13
Government/surveys too invasive	4	1.78	18	7.00	40	8.49	62	7.85
Gatekeeper/household member won't allow participation	5	1.77	58	19.04	32	6.61	95	7.54
Confidentiality or survey legitimacy concerns	1	0.39	10	2.94	26	4.93	37	4.40
House too messy/too ill	0	0.00	0	0.00	6	1.42	6	1.17
Other	0	0.00	4	1.04	3	0.40	7	0.43
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (North Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	39	100.00	35	100.00	72	100.00	146	100.00
Parental refusal	36	92.31	0	0.00	0	0.00	36	24.66
Nothing in it for me	1	2.56	18	51.43	24	33.33	43	29.45
No time	0	0.00	8	22.86	20	27.78	28	19.18
Government/surveys too invasive	0	0.00	2	5.71	11	15.28	13	8.90
Gatekeeper/household member won't allow participation	1	2.56	6	17.14	15	20.83	22	15.07
Confidentiality or survey legitimacy concerns	0	0.00	1	2.86	1	1.39	2	1.37
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	1	2.56	0	0.00	1	1.39	2	1.37
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (North Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	39	100.00	35	100.00	72	100.00	146	100.00
Parental refusal	36	91.29	0	0.00	0	0.00	36	5.93
Nothing in it for me	1	1.78	18	44.92	24	31.76	43	30.74
No time	0	0.00	8	23.30	20	26.91	28	24.91
Government/surveys too invasive	0	0.00	2	3.43	11	16.80	13	14.76
Gatekeeper/household member won't allow participation	1	1.77	6	26.08	15	21.82	22	20.82
Confidentiality or survey legitimacy concerns	0	0.00	1	2.27	1	0.86	2	0.90
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	1	5.16	0	0.00	1	1.85	2	1.94
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (North Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	60	100.00	61	100.00	116	100.00	237	100.00
Parental refusal	33	55.00	0	0.00	0	0.00	33	13.92
Nothing in it for me	5	8.33	20	32.79	18	15.52	43	18.14
No time	8	13.33	34	55.74	47	40.52	89	37.55
Government/surveys too invasive	8	13.33	5	8.20	43	37.07	56	23.63
Gatekeeper/household member won't allow participation	5	8.33	1	1.64	5	4.31	11	4.64
Confidentiality or survey legitimacy concerns	0	0.00	1	1.64	1	0.86	2	0.84
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	1	1.67	0	0.00	2	1.72	3	1.27
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (North Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	60	100.00	61	100.00	116	100.00	237	100.00
Parental refusal	33	58.77	0	0.00	0	0.00	33	3.25
Nothing in it for me	5	7.36	20	32.83	18	15.20	43	16.80
No time	8	12.78	34	54.93	47	38.55	89	39.01
Government/surveys too invasive	8	11.58	5	8.12	43	40.80	56	35.42
Gatekeeper/household member won't allow participation	5	8.07	1	1.72	5	3.59	11	3.63
Confidentiality or survey legitimacy concerns	0	0.00	1	2.40	1	0.66	2	0.83
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	1	1.44	0	0.00	2	1.19	3	1.07
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Ohio) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	258	100.00	221	100.00	376	100.00	855	100.00
Parental refusal	199	77.13	0	0.00	0	0.00	199	23.27
Nothing in it for me	31	12.02	102	46.15	189	50.27	322	37.66
No time	10	3.88	73	33.03	109	28.99	192	22.46
Government/surveys too invasive	5	1.94	13	5.88	45	11.97	63	7.37
Gatekeeper/household member won't allow participation	11	4.26	24	10.86	11	2.93	46	5.38
Confidentiality or survey legitimacy concerns	1	0.39	4	1.81	18	4.79	23	2.69
House too messy/too ill	0	0.00	0	0.00	4	1.06	4	0.47
Other	0	0.00	4	1.81	0	0.00	4	0.47
Missing	1	0.39	1	0.45	0	0.00	2	0.23

Table 7.25 2012 Interview Refusal Reasons, by Age (Ohio) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	258	100.00	221	100.00	376	100.00	855	100.00
Parental refusal	199	77.34	0	0.00	0	0.00	199	5.74
Nothing in it for me	31	11.97	102	45.89	189	52.79	322	49.13
No time	10	3.61	73	33.47	109	25.37	192	24.50
Government/surveys too invasive	5	2.01	13	5.83	45	12.33	63	10.97
Gatekeeper/household member won't allow participation	11	4.34	24	10.64	11	2.95	46	3.75
Confidentiality or survey legitimacy concerns	1	0.36	4	2.09	18	5.30	23	4.64
House too messy/too ill	0	0.00	0	0.00	4	1.26	4	1.05
Other	0	0.00	4	1.69	0	0.00	4	0.16
Missing	1	0.36	1	0.40	0	0.00	2	0.06

Table 7.24 2012 Interview Refusal Reasons, by Age (Oklahoma) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	71	100.00	71	100.00	96	100.00	238	100.00
Parental refusal	58	81.69	0	0.00	0	0.00	58	24.37
Nothing in it for me	9	12.68	54	76.06	62	64.58	125	52.52
No time	4	5.63	10	14.08	17	17.71	31	13.03
Government/surveys too invasive	0	0.00	2	2.82	7	7.29	9	3.78
Gatekeeper/household member won't allow participation	0	0.00	4	5.63	4	4.17	8	3.36
Confidentiality or survey legitimacy concerns	0	0.00	1	1.41	2	2.08	3	1.26
House too messy/too ill	0	0.00	0	0.00	3	3.13	3	1.26
Other	0	0.00	0	0.00	1	1.04	1	0.42
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Oklahoma) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	71	100.00	71	100.00	96	100.00	238	100.00
Parental refusal	58	81.51	0	0.00	0	0.00	58	6.53
Nothing in it for me	9	13.83	54	74.05	62	62.27	125	59.75
No time	4	4.67	10	12.72	17	17.36	31	15.81
Government/surveys too invasive	0	0.00	2	2.55	7	8.92	9	7.47
Gatekeeper/household member won't allow participation	0	0.00	4	8.75	4	4.37	8	4.52
Confidentiality or survey legitimacy concerns	0	0.00	1	1.92	2	2.62	3	2.33
House too messy/too ill	0	0.00	0	0.00	3	3.38	3	2.72
Other	0	0.00	0	0.00	1	1.08	1	0.87
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Oregon) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	30	100.00	55	100.00	82	100.00	167	100.00
Parental refusal	17	56.67	0	0.00	0	0.00	17	10.18
Nothing in it for me	6	20.00	29	52.73	43	52.44	78	46.71
No time	3	10.00	17	30.91	27	32.93	47	28.14
Government/surveys too invasive	3	10.00	3	5.45	4	4.88	10	5.99
Gatekeeper/household member won't allow participation	0	0.00	3	5.45	1	1.22	4	2.40
Confidentiality or survey legitimacy concerns	0	0.00	2	3.64	3	3.66	5	2.99
House too messy/too ill	0	0.00	0	0.00	1	1.22	1	0.60
Other	1	3.33	1	1.82	3	3.66	5	2.99
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Oregon) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	30	100.00	55	100.00	82	100.00	167	100.00
Parental refusal	17	52.11	0	0.00	0	0.00	17	2.42
Nothing in it for me	6	18.55	29	54.18	43	48.29	78	47.44
No time	3	8.28	17	30.05	27	32.63	47	31.27
Government/surveys too invasive	3	17.43	3	3.75	4	5.65	10	6.03
Gatekeeper/household member won't allow participation	0	0.00	3	5.52	1	1.82	4	2.07
Confidentiality or survey legitimacy concerns	0	0.00	2	4.20	3	6.78	5	6.24
House too messy/too ill	0	0.00	0	0.00	1	0.84	1	0.73
Other	1	3.62	1	2.31	3	3.98	5	3.82
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Pennsylvania) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	208	100.00	229	100.00	447	100.00	884	100.00
Parental refusal	149	71.63	0	0.00	0	0.00	149	16.86
Nothing in it for me	30	14.42	92	40.17	190	42.51	312	35.29
No time	12	5.77	59	25.76	111	24.83	182	20.59
Government/surveys too invasive	7	3.37	16	6.99	61	13.65	84	9.50
Gatekeeper/household member won't allow participation	6	2.88	33	14.41	11	2.46	50	5.66
Confidentiality or survey legitimacy concerns	3	1.44	23	10.04	60	13.42	86	9.73
House too messy/too ill	0	0.00	0	0.00	10	2.24	10	1.13
Other	1	0.48	6	2.62	4	0.89	11	1.24
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Pennsylvania) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	208	100.00	229	100.00	447	100.00	884	100.00
Parental refusal	149	73.41	0	0.00	0	0.00	149	4.08
Nothing in it for me	30	13.31	92	42.21	190	43.18	312	41.44
No time	12	5.14	59	23.68	111	24.00	182	22.93
Government/surveys too invasive	7	3.84	16	8.79	61	13.22	84	12.34
Gatekeeper/household member won't allow participation	6	2.79	33	14.41	11	2.35	50	3.35
Confidentiality or survey legitimacy concerns	3	0.97	23	8.66	60	13.16	86	12.12
House too messy/too ill	0	0.00	0	0.00	10	3.29	10	2.84
Other	1	0.53	6	2.25	4	0.80	11	0.90
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Rhode Island) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	43	100.00	85	100.00	165	100.00
Parental refusal	28	75.68	0	0.00	0	0.00	28	16.97
Nothing in it for me	5	13.51	14	32.56	23	27.06	42	25.45
No time	2	5.41	10	23.26	31	36.47	43	26.06
Government/surveys too invasive	0	0.00	3	6.98	16	18.82	19	11.52
Gatekeeper/household member won't allow participation	1	2.70	12	27.91	7	8.24	20	12.12
Confidentiality or survey legitimacy concerns	1	2.70	4	9.30	7	8.24	12	7.27
House too messy/too ill	0	0.00	0	0.00	1	1.18	1	0.61
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Rhode Island) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	37	100.00	43	100.00	85	100.00	165	100.00
Parental refusal	28	74.58	0	0.00	0	0.00	28	4.26
Nothing in it for me	5	14.56	14	25.97	23	24.94	42	24.44
No time	2	5.60	10	25.89	31	34.05	43	31.70
Government/surveys too invasive	0	0.00	3	6.99	16	18.78	19	16.67
Gatekeeper/household member won't allow participation	1	2.29	12	28.80	7	8.83	20	10.22
Confidentiality or survey legitimacy concerns	1	2.97	4	12.34	7	11.49	12	11.08
House too messy/too ill	0	0.00	0	0.00	1	1.91	1	1.63
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (South Carolina) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	54	100.00	27	100.00	71	100.00	152	100.00
Parental refusal	53	98.15	0	0.00	0	0.00	53	34.87
Nothing in it for me	1	1.85	11	40.74	27	38.03	39	25.66
No time	0	0.00	5	18.52	17	23.94	22	14.47
Government/surveys too invasive	0	0.00	5	18.52	12	16.90	17	11.18
Gatekeeper/household member won't allow participation	0	0.00	2	7.41	6	8.45	8	5.26
Confidentiality or survey legitimacy concerns	0	0.00	4	14.81	5	7.04	9	5.92
House too messy/too ill	0	0.00	0	0.00	2	2.82	2	1.32
Other	0	0.00	0	0.00	2	2.82	2	1.32
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (South Carolina) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	54	100.00	27	100.00	71	100.00	152	100.00
Parental refusal	53	98.20	0	0.00	0	0.00	53	8.34
Nothing in it for me	1	1.80	11	43.03	27	38.27	39	35.47
No time	0	0.00	5	18.81	17	24.33	22	21.92
Government/surveys too invasive	0	0.00	5	17.68	12	15.53	17	14.34
Gatekeeper/household member won't allow participation	0	0.00	2	6.74	6	6.91	8	6.31
Confidentiality or survey legitimacy concerns	0	0.00	4	13.74	5	6.79	9	6.65
House too messy/too ill	0	0.00	0	0.00	2	3.88	2	3.31
Other	0	0.00	0	0.00	2	4.29	2	3.65
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (South Dakota) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	40	100.00	37	100.00	82	100.00	159	100.00
Parental refusal	29	72.50	0	0.00	0	0.00	29	18.24
Nothing in it for me	4	10.00	27	72.97	39	47.56	70	44.03
No time	3	7.50	4	10.81	23	28.05	30	18.87
Government/surveys too invasive	3	7.50	3	8.11	13	15.85	19	11.95
Gatekeeper/household member won't allow participation	1	2.50	2	5.41	3	3.66	6	3.77
Confidentiality or survey legitimacy concerns	0	0.00	1	2.70	4	4.88	5	3.14
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (South Dakota) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	40	100.00	37	100.00	82	100.00	159	100.00
Parental refusal	29	71.77	0	0.00	0	0.00	29	4.32
Nothing in it for me	4	9.70	27	71.26	39	43.64	70	43.45
No time	3	8.16	4	11.28	23	30.59	30	27.94
Government/surveys too invasive	3	8.07	3	9.11	13	17.11	19	16.03
Gatekeeper/household member won't allow participation	1	2.30	2	5.57	3	3.66	6	3.71
Confidentiality or survey legitimacy concerns	0	0.00	1	2.78	4	4.99	5	4.54
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Tennessee) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	22	100.00	36	100.00	54	100.00	112	100.00
Parental refusal	19	86.36	0	0.00	0	0.00	19	16.96
Nothing in it for me	2	9.09	18	50.00	25	46.30	45	40.18
No time	0	0.00	11	30.56	13	24.07	24	21.43
Government/surveys too invasive	0	0.00	0	0.00	2	3.70	2	1.79
Gatekeeper/household member won't allow participation	1	4.55	6	16.67	9	16.67	16	14.29
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.85	1	0.89
House too messy/too ill	0	0.00	1	2.78	2	3.70	3	2.68
Other	0	0.00	0	0.00	2	3.70	2	1.79
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Tennessee) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	22	100.00	36	100.00	54	100.00	112	100.00
Parental refusal	19	90.57	0	0.00	0	0.00	19	5.42
Nothing in it for me	2	6.45	18	56.44	25	44.84	45	43.56
No time	0	0.00	11	25.12	13	25.44	24	23.89
Government/surveys too invasive	0	0.00	0	0.00	2	4.15	2	3.54
Gatekeeper/household member won't allow participation	1	2.98	6	16.37	9	18.79	16	17.63
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	1	1.12	1	0.95
House too messy/too ill	0	0.00	1	2.06	2	3.38	3	3.07
Other	0	0.00	0	0.00	2	2.28	2	1.94
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Texas) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	156	100.00	145	100.00	305	100.00	606	100.00
Parental refusal	126	80.77	0	0.00	0	0.00	126	20.79
Nothing in it for me	13	8.33	79	54.48	149	48.85	241	39.77
No time	9	5.77	19	13.10	57	18.69	85	14.03
Government/surveys too invasive	2	1.28	10	6.90	43	14.10	55	9.08
Gatekeeper/household member won't allow participation	4	2.56	26	17.93	24	7.87	54	8.91
Confidentiality or survey legitimacy concerns	0	0.00	6	4.14	15	4.92	21	3.47
House too messy/too ill	0	0.00	0	0.00	2	0.66	2	0.33
Other	2	1.28	5	3.45	15	4.92	22	3.63
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Texas) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	156	100.00	145	100.00	305	100.00	606	100.00
Parental refusal	126	81.04	0	0.00	0	0.00	126	5.72
Nothing in it for me	13	7.38	79	51.61	149	46.61	241	44.26
No time	9	4.59	19	14.23	57	19.55	85	18.04
Government/surveys too invasive	2	2.53	10	9.48	43	15.12	55	13.75
Gatekeeper/household member won't allow participation	4	2.86	26	17.29	24	6.74	54	7.36
Confidentiality or survey legitimacy concerns	0	0.00	6	4.05	15	6.40	21	5.75
House too messy/too ill	0	0.00	0	0.00	2	0.84	2	0.71
Other	2	1.60	5	3.33	15	4.76	22	4.42
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Utah) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	23	100.00	46	100.00	44	100.00	113	100.00
Parental refusal	16	69.57	0	0.00	0	0.00	16	14.16
Nothing in it for me	4	17.39	32	69.57	24	54.55	60	53.10
No time	2	8.70	7	15.22	5	11.36	14	12.39
Government/surveys too invasive	0	0.00	0	0.00	10	22.73	10	8.85
Gatekeeper/household member won't allow participation	1	4.35	4	8.70	5	11.36	10	8.85
Confidentiality or survey legitimacy concerns	0	0.00	1	2.17	0	0.00	1	0.88
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	2	4.35	0	0.00	2	1.77

Table 7.25 2012 Interview Refusal Reasons, by Age (Utah) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	23	100.00	46	100.00	44	100.00	113	100.00
Parental refusal	16	76.64	0	0.00	0	0.00	16	5.54
Nothing in it for me	4	14.11	32	71.35	24	56.68	60	56.18
No time	2	5.43	7	12.29	5	9.84	14	9.95
Government/surveys too invasive	0	0.00	0	0.00	10	23.46	10	17.64
Gatekeeper/household member won't allow participation	1	3.82	4	9.62	5	10.02	10	9.50
Confidentiality or survey legitimacy concerns	0	0.00	1	1.97	0	0.00	1	0.35
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	2	4.77	0	0.00	2	0.84

Table 7.24 2012 Interview Refusal Reasons, by Age (Vermont) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	50	100.00	80	100.00	177	100.00
Parental refusal	28	59.57	0	0.00	0	0.00	28	15.82
Nothing in it for me	8	17.02	11	22.00	27	33.75	46	25.99
No time	8	17.02	29	58.00	34	42.50	71	40.11
Government/surveys too invasive	3	6.38	8	16.00	16	20.00	27	15.25
Gatekeeper/household member won't allow participation	0	0.00	2	4.00	1	1.25	3	1.69
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	1	1.25	1	0.56
Other	0	0.00	0	0.00	1	1.25	1	0.56
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Vermont) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	47	100.00	50	100.00	80	100.00	177	100.00
Parental refusal	28	65.30	0	0.00	0	0.00	28	4.09
Nothing in it for me	8	15.13	11	22.12	27	34.07	46	31.89
No time	8	14.01	29	57.49	34	36.89	71	37.17
Government/surveys too invasive	3	5.56	8	14.10	16	26.14	27	23.85
Gatekeeper/household member won't allow participation	0	0.00	2	6.29	1	0.69	3	1.12
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	1	1.54	1	1.31
Other	0	0.00	0	0.00	1	0.67	1	0.57
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	36	100.00	36	100.00	81	100.00	153	100.00
Parental refusal	27	75.00	0	0.00	0	0.00	27	17.65
Nothing in it for me	8	22.22	18	50.00	52	64.20	78	50.98
No time	0	0.00	7	19.44	21	25.93	28	18.30
Government/surveys too invasive	1	2.78	0	0.00	3	3.70	4	2.61
Gatekeeper/household member won't allow participation	0	0.00	9	25.00	1	1.23	10	6.54
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	2.47	2	1.31
House too messy/too ill	0	0.00	0	0.00	2	2.47	2	1.31
Other	0	0.00	2	5.56	0	0.00	2	1.31
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	36	100.00	36	100.00	81	100.00	153	100.00
Parental refusal	27	77.68	0	0.00	0	0.00	27	4.10
Nothing in it for me	8	20.00	18	46.93	52	65.28	78	61.46
No time	0	0.00	7	18.15	21	24.89	28	23.05
Government/surveys too invasive	1	2.32	0	0.00	3	4.15	4	3.73
Gatekeeper/household member won't allow participation	0	0.00	9	28.70	1	0.80	10	2.94
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	2	2.36	2	2.05
House too messy/too ill	0	0.00	0	0.00	2	2.52	2	2.19
Other	0	0.00	2	6.23	0	0.00	2	0.49
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Washington) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	55	100.00	63	100.00	93	100.00	211	100.00
Parental refusal	42	76.36	0	0.00	0	0.00	42	19.91
Nothing in it for me	8	14.55	33	52.38	46	49.46	87	41.23
No time	3	5.45	14	22.22	28	30.11	45	21.33
Government/surveys too invasive	2	3.64	2	3.17	9	9.68	13	6.16
Gatekeeper/household member won't allow participation	0	0.00	11	17.46	2	2.15	13	6.16
Confidentiality or survey legitimacy concerns	0	0.00	3	4.76	8	8.60	11	5.21
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Washington) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	55	100.00	63	100.00	93	100.00	211	100.00
Parental refusal	42	72.94	0	0.00	0	0.00	42	4.95
Nothing in it for me	8	16.99	33	49.71	46	48.49	87	46.48
No time	3	5.97	14	21.64	28	31.14	45	28.47
Government/surveys too invasive	2	4.10	2	5.81	9	11.20	13	10.17
Gatekeeper/household member won't allow participation	0	0.00	11	17.69	2	1.48	13	3.03
Confidentiality or survey legitimacy concerns	0	0.00	3	5.15	8	7.69	11	6.91
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (West Virginia) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	29	100.00	45	100.00	89	100.00	163	100.00
Parental refusal	17	58.62	0	0.00	0	0.00	17	10.43
Nothing in it for me	7	24.14	24	53.33	43	48.31	74	45.40
No time	0	0.00	13	28.89	22	24.72	35	21.47
Government/surveys too invasive	4	13.79	2	4.44	12	13.48	18	11.04
Gatekeeper/household member won't allow participation	1	3.45	5	11.11	2	2.25	8	4.91
Confidentiality or survey legitimacy concerns	0	0.00	1	2.22	6	6.74	7	4.29
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	4	4.49	4	2.45
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (West Virginia) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	29	100.00	45	100.00	89	100.00	163	100.00
Parental refusal	17	56.45	0	0.00	0	0.00	17	2.04
Nothing in it for me	7	26.06	24	53.65	43	47.24	74	46.86
No time	0	0.00	13	27.38	22	24.67	35	23.94
Government/surveys too invasive	4	14.46	2	5.60	12	14.14	18	13.63
Gatekeeper/household member won't allow participation	1	3.04	5	11.52	2	2.71	8	3.26
Confidentiality or survey legitimacy concerns	0	0.00	1	1.86	6	6.14	7	5.66
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	4	5.10	4	4.61
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Wisconsin) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	71	100.00	68	100.00	174	100.00
Parental refusal	18	51.43	0	0.00	0	0.00	18	10.34
Nothing in it for me	13	37.14	40	56.34	40	58.82	93	53.45
No time	2	5.71	16	22.54	17	25.00	35	20.11
Government/surveys too invasive	1	2.86	6	8.45	6	8.82	13	7.47
Gatekeeper/household member won't allow participation	0	0.00	7	9.86	3	4.41	10	5.75
Confidentiality or survey legitimacy concerns	1	2.86	2	2.82	2	2.94	5	2.87
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Wisconsin) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	35	100.00	71	100.00	68	100.00	174	100.00
Parental refusal	18	54.32	0	0.00	0	0.00	18	3.37
Nothing in it for me	13	35.45	40	55.50	40	53.58	93	52.69
No time	2	5.96	16	24.50	17	28.48	35	26.59
Government/surveys too invasive	1	0.71	6	8.76	6	10.70	13	9.84
Gatekeeper/household member won't allow participation	0	0.00	7	8.88	3	5.33	10	5.44
Confidentiality or survey legitimacy concerns	1	3.56	2	2.36	2	1.91	5	2.07
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.24 2012 Interview Refusal Reasons, by Age (Wyoming) (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	28	100.00	41	100.00	71	100.00	140	100.00
Parental refusal	20	71.43	0	0.00	0	0.00	20	14.29
Nothing in it for me	4	14.29	11	26.83	20	28.17	35	25.00
No time	1	3.57	14	34.15	22	30.99	37	26.43
Government/surveys too invasive	2	7.14	4	9.76	15	21.13	21	15.00
Gatekeeper/household member won't allow participation	1	3.57	3	7.32	2	2.82	6	4.29
Confidentiality or survey legitimacy concerns	0	0.00	3	7.32	7	9.86	10	7.14
House too messy/too ill	0	0.00	0	0.00	1	1.41	1	0.71
Other	0	0.00	6	14.63	4	5.63	10	7.14
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25 2012 Interview Refusal Reasons, by Age (Wyoming) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Refusal Cases	28	100.00	41	100.00	71	100.00	140	100.00
Parental refusal	20	75.48	0	0.00	0	0.00	20	4.20
Nothing in it for me	4	13.16	11	34.62	20	29.03	35	28.70
No time	1	2.70	14	29.72	22	29.81	37	28.29
Government/surveys too invasive	2	5.07	4	8.14	15	21.05	21	18.89
Gatekeeper/household member won't allow participation	1	3.58	3	7.55	2	3.17	6	3.62
Confidentiality or survey legitimacy concerns	0	0.00	3	8.21	7	9.89	10	9.17
House too messy/too ill	0	0.00	0	0.00	1	1.99	1	1.69
Other	0	0.00	6	11.75	4	5.06	10	5.43
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25a 2012 Interview Refusal Reasons, by Age and Race/Ethnicity (Total United States) (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Hispanic or Latino								
Refusal Cases	451	100.00	619	100.00	693	100.00	1,763	100.00
Parental refusal	326	68.19	0	0.00	0	0.00	326	5.38
Nothing in it for me	72	15.43	353	55.51	377	54.15	802	51.28
No time	21	4.11	126	19.72	167	23.57	314	21.51
Government/surveys too invasive	9	1.91	23	2.98	42	5.62	74	4.97
Gatekeeper/household member won't allow participation	12	3.87	61	11.35	32	5.49	105	6.16
Confidentiality or survey legitimacy concerns	6	5.77	20	2.53	38	5.62	64	5.21
House too messy/too ill	1	0.17	0	0.00	2	0.45	3	0.36
Other	4	0.55	35	7.79	35	5.11	74	5.12
Missing	0	0.00	1	0.12	0	0.00	1	0.02
Not Hispanic or Latino								
Black or African American								
Refusal Cases	301	100.00	310	100.00	461	100.00	1,072	100.00
Parental refusal	215	71.44	0	0.00	0	0.00	215	5.25
Nothing in it for me	53	16.96	171	54.27	281	60.98	505	57.06
No time	9	3.57	58	15.57	90	18.65	157	17.23
Government/surveys too invasive	11	4.50	15	5.00	28	7.02	54	6.62
Gatekeeper/household member won't allow participation	12	3.46	52	20.66	25	5.06	89	6.54
Confidentiality or survey legitimacy concerns	1	0.07	11	3.95	27	6.55	39	5.80
House too messy/too ill	0	0.00	0	0.00	5	0.94	5	0.77
Other	0	0.00	3	0.54	4	0.63	7	0.57
Missing	0	0.00	0	0.00	1	0.17	1	0.14

Table 7.25a 2012 Interview Refusal Reasons, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino White								
Refusal Cases	2,585	100.00	2,722	100.00	5,019	100.00	10,326	100.00
Parental refusal	1,966	76.78	0	0.00	0	0.00	1,966	4.53
Nothing in it for me	325	12.67	1,291	49.11	2,347	47.73	3,963	45.78
No time	134	4.56	768	25.58	1,431	25.99	2,333	24.69
Government/surveys too invasive	69	2.55	170	6.30	635	12.80	874	11.65
Gatekeeper/household member won't allow participation	67	2.11	372	14.76	220	4.85	659	5.52
Confidentiality or survey legitimacy concerns	13	0.68	86	3.03	261	5.93	360	5.37
House too messy/too ill	0	0.00	5	0.50	65	1.56	70	1.38
Other	10	0.62	25	0.57	54	1.06	89	0.99
Missing	1	0.02	5	0.15	6	0.09	12	0.09
Not Hispanic or Latino American Indian or Alaska Native								
Refusal Cases	35	100.00	35	100.00	46	100.00	116	100.00
Parental refusal	27	80.94	0	0.00	0	0.00	27	7.51
Nothing in it for me	4	6.28	18	55.78	25	55.08	47	50.68
No time	2	5.18	8	17.41	14	32.67	24	27.27
Government/surveys too invasive	1	1.28	2	11.19	4	7.03	7	7.27
Gatekeeper/household member won't allow participation	1	6.33	3	2.52	2	3.88	6	3.85
Confidentiality or survey legitimacy concerns	0	0.00	4	13.11	0	0.00	4	2.45
House too messy/too ill	0	0.00	0	0.00	1	1.34	1	0.97
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.25a 2012 Interview Refusal Reasons, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino								
Native Hawaiian or Other Pacific Islander								
Refusal Cases	11	100.00	13	100.00	22	100.00	46	100.00
Parental refusal	9	95.33	0	0.00	0	0.00	9	4.01
Nothing in it for me	1	3.06	6	48.12	11	56.98	18	53.74
No time	0	0.00	4	40.05	11	43.02	15	40.89
Government/surveys too invasive	0	0.00	1	2.14	0	0.00	1	0.23
Gatekeeper/household member won't allow participation	1	1.61	0	0.00	0	0.00	1	0.07
Confidentiality or survey legitimacy concerns	0	0.00	0	0.00	0	0.00	0	0.00
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	1	7.99	0	0.00	1	0.88
Missing	0	0.00	1	1.70	0	0.00	1	0.19
Not Hispanic or Latino								
Asian								
Refusal Cases	157	100.00	178	100.00	297	100.00	632	100.00
Parental refusal	125	74.26	0	0.00	0	0.00	125	6.58
Nothing in it for me	15	12.54	81	42.77	152	53.21	248	48.39
No time	4	2.19	39	21.45	80	24.48	123	22.15
Government/surveys too invasive	5	5.39	15	10.16	27	9.85	47	9.49
Gatekeeper/household member won't allow participation	5	3.30	28	15.94	18	6.95	51	7.68
Confidentiality or survey legitimacy concerns	2	2.20	13	9.42	13	4.56	28	4.92
House too messy/too ill	0	0.00	0	0.00	3	0.55	3	0.44
Other	1	0.12	2	0.25	2	0.08	5	0.10
Missing	0	0.00	0	0.00	2	0.32	2	0.26

Table 7.25a 2012 Interview Refusal Reasons, by Age and Race/Ethnicity (Total United States) (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Not Hispanic or Latino Multiple Races								
Refusal Cases	147	100.00	99	100.00	74	100.00	320	100.00
Parental refusal	119	78.04	0	0.00	0	0.00	119	16.83
Nothing in it for me	17	9.79	44	46.62	41	57.08	102	44.95
No time	6	4.52	25	22.45	17	19.09	48	16.57
Government/surveys too invasive	4	6.90	8	8.06	7	8.10	19	7.83
Gatekeeper/household member won't allow participation	1	0.74	18	18.28	5	7.97	24	8.32
Confidentiality or survey legitimacy concerns	0	0.00	4	4.59	4	7.76	8	5.50
House too messy/too ill	0	0.00	0	0.00	0	0.00	0	0.00
Other	0	0.00	0	0.00	0	0.00	0	0.00
Missing	0	0.00	0	0.00	0	0.00	0	0.00

Table 7.26 2012 Interview Results, by Small Age Groups and Gender (Unweighted Percentages)

	Male		Female		Total	
	Count	%	Count	%	Count	%
12-13						
Eligible Cases	4,605	100.00	4,419	100.00	9,024	100.00
70 - Interview Complete	3,751	81.45	3,664	82.91	7,415	82.17
71 - No One at DU*	70	1.52	70	1.58	140	1.55
77 - Refusal	107	2.32	107	2.42	214	2.37
Other	677	14.70	578	13.08	1,255	13.91
14-15						
Eligible Cases	4,542	100.00	4,350	100.00	8,892	100.00
70 - Interview Complete	3,800	83.66	3,647	83.84	7,447	83.75
71 - No One at DU*	84	1.85	76	1.75	160	1.80
77 - Refusal	138	3.04	119	2.74	257	2.89
Other	520	11.45	508	11.68	1,028	11.56
16-17						
Eligible Cases	4,659	100.00	4,572	100.00	9,231	100.00
70 - Interview Complete	3,843	82.49	3,787	82.83	7,630	82.66
71 - No One at DU*	85	1.82	83	1.82	168	1.82
77 - Refusal	218	4.68	211	4.62	429	4.65
Other	513	11.01	491	10.74	1,004	10.88
18-20						
Eligible Cases	5,117	100.00	5,127	100.00	10,244	100.00
70 - Interview Complete	4,158	81.26	4,262	83.13	8,420	82.19
71 - No One at DU*	167	3.26	146	2.85	313	3.06
77 - Refusal	644	12.59	583	11.37	1,227	11.98
Other	148	2.89	136	2.65	284	2.77

Table 7.26 2012 Interview Results, by Small Age Groups and Gender (Unweighted Percentages) (continued)

	Male		Female		Total	
	Count	%	Count	%	Count	%
21-25						
Eligible Cases	8,830	100.00	9,565	100.00	18,395	100.00
70 - Interview Complete	6,733	76.25	7,609	79.55	14,342	77.97
71 - No One at DU*	369	4.18	408	4.27	777	4.22
77 - Refusal	1,440	16.31	1,309	13.69	2,749	14.94
Other	288	3.26	239	2.50	527	2.86
26-29						
Eligible Cases	1,844	100.00	1,886	100.00	3,730	100.00
70 - Interview Complete	1,353	73.37	1,473	78.10	2,826	75.76
71 - No One at DU*	92	4.99	68	3.61	160	4.29
77 - Refusal	349	18.93	301	15.96	650	17.43
Other	50	2.71	44	2.33	94	2.52
30-34						
Eligible Cases	2,165	100.00	2,409	100.00	4,574	100.00
70 - Interview Complete	1,571	72.56	1,891	78.50	3,462	75.69
71 - No One at DU*	100	4.62	71	2.95	171	3.74
77 - Refusal	430	19.86	393	16.31	823	17.99
Other	64	2.96	54	2.24	118	2.58
35-39						
Eligible Cases	1,933	100.00	1,975	100.00	3,908	100.00
70 - Interview Complete	1,364	70.56	1,522	77.06	2,886	73.85
71 - No One at DU*	100	5.17	62	3.14	162	4.15
77 - Refusal	435	22.50	340	17.22	775	19.83
Other	34	1.76	51	2.58	85	2.18

Table 7.26 2012 Interview Results, by Small Age Groups and Gender (Unweighted Percentages) (continued)

	Male		Female		Total	
	Count	%	Count	%	Count	%
40-44						
Eligible Cases	1,983	100.00	2,133	100.00	4,116	100.00
70 - Interview Complete	1,392	70.20	1,625	76.18	3,017	73.30
71 - No One at DU*	81	4.08	70	3.28	151	3.67
77 - Refusal	461	23.25	384	18.00	845	20.53
Other	49	2.47	54	2.53	103	2.50
45-49						
Eligible Cases	2,061	100.00	2,279	100.00	4,340	100.00
70 - Interview Complete	1,416	68.70	1,745	76.57	3,161	72.83
71 - No One at DU*	83	4.03	62	2.72	145	3.34
77 - Refusal	506	24.55	423	18.56	929	21.41
Other	56	2.72	49	2.15	105	2.42
50+						
Eligible Cases	5,203	100.00	5,999	100.00	11,202	100.00
70 - Interview Complete	3,488	67.04	4,215	70.26	7,703	68.76
71 - No One at DU*	135	2.59	110	1.83	245	2.19
77 - Refusal	1,276	24.52	1,314	21.90	2,590	23.12
Other	304	5.84	360	6.00	664	5.93
Total						
Eligible Cases	42,942	100.00	44,714	100.00	87,656	100.00
70 - Interview Complete	32,869	76.54	35,440	79.26	68,309	77.93
71 - No One at DU*	1,366	3.18	1,226	2.74	2,592	2.96
77 - Refusal	6,004	13.98	5,484	12.26	11,488	13.11
Other	2,703	6.29	2,564	5.73	5,267	6.01

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.27 2012 Interview Results, by Small Age Groups and Gender (Weighted Percentages)

	Male		Female		Total	
	Count	%	Count	%	Count	%
12-13						
Eligible Cases	4,605	100.00	4,419	100.00	9,024	100.00
70 - Interview Complete	3,751	81.39	3,664	83.42	7,415	82.39
71 - No One at DU*	70	1.48	70	1.53	140	1.50
77 - Refusal	107	2.76	107	2.53	214	2.65
Other	677	14.37	578	12.51	1,255	13.46
14-15						
Eligible Cases	4,542	100.00	4,350	100.00	8,892	100.00
70 - Interview Complete	3,800	83.14	3,647	83.50	7,447	83.32
71 - No One at DU*	84	1.95	76	1.84	160	1.90
77 - Refusal	138	2.84	119	2.74	257	2.79
Other	520	12.06	508	11.93	1,028	12.00
16-17						
Eligible Cases	4,659	100.00	4,572	100.00	9,231	100.00
70 - Interview Complete	3,843	82.34	3,787	83.26	7,630	82.79
71 - No One at DU*	85	1.71	83	1.54	168	1.63
77 - Refusal	218	4.61	211	4.63	429	4.62
Other	513	11.34	491	10.56	1,004	10.96
18-20						
Eligible Cases	5,117	100.00	5,127	100.00	10,244	100.00
70 - Interview Complete	4,158	81.29	4,262	82.92	8,420	82.09
71 - No One at DU*	167	3.11	146	2.90	313	3.01
77 - Refusal	644	12.45	583	11.43	1,227	11.95
Other	148	3.15	136	2.74	284	2.95

Table 7.27 2012 Interview Results, by Small Age Groups and Gender (Weighted Percentages) (continued)

	Male		Female		Total	
	Count	%	Count	%	Count	%
21-25						
Eligible Cases	8,830	100.00	9,565	100.00	18,395	100.00
70 - Interview Complete	6,733	76.08	7,609	79.15	14,342	77.64
71 - No One at DU*	369	3.77	408	4.10	777	3.94
77 - Refusal	1,440	16.74	1,309	14.01	2,749	15.35
Other	288	3.42	239	2.74	527	3.07
26-29						
Eligible Cases	1,844	100.00	1,886	100.00	3,730	100.00
70 - Interview Complete	1,353	73.45	1,473	78.21	2,826	75.79
71 - No One at DU*	92	5.26	68	3.96	160	4.62
77 - Refusal	349	18.08	301	14.74	650	16.44
Other	50	3.21	44	3.09	94	3.15
30-34						
Eligible Cases	2,165	100.00	2,409	100.00	4,574	100.00
70 - Interview Complete	1,571	71.36	1,891	78.15	3,462	74.96
71 - No One at DU*	100	5.20	71	2.86	171	3.95
77 - Refusal	430	19.74	393	16.47	823	18.00
Other	64	3.71	54	2.53	118	3.08
35-39						
Eligible Cases	1,933	100.00	1,975	100.00	3,908	100.00
70 - Interview Complete	1,364	69.66	1,522	77.16	2,886	73.34
71 - No One at DU*	100	5.81	62	2.80	162	4.33
77 - Refusal	435	22.78	340	17.44	775	20.16
Other	34	1.75	51	2.60	85	2.17

Table 7.27 2012 Interview Results, by Small Age Groups and Gender (Weighted Percentages) (continued)

	Male		Female		Total	
	Count	%	Count	%	Count	%
40-44						
Eligible Cases	1,983	100.00	2,133	100.00	4,116	100.00
70 - Interview Complete	1,392	69.49	1,625	76.74	3,017	73.27
71 - No One at DU*	81	4.43	70	2.99	151	3.68
77 - Refusal	461	23.57	384	17.53	845	20.42
Other	49	2.51	54	2.74	103	2.63
45-49						
Eligible Cases	2,061	100.00	2,279	100.00	4,340	100.00
70 - Interview Complete	1,416	68.71	1,745	75.92	3,161	72.44
71 - No One at DU*	83	3.92	62	2.62	145	3.25
77 - Refusal	506	24.25	423	18.52	929	21.29
Other	56	3.12	49	2.93	105	3.02
50+						
Eligible Cases	5,203	100.00	5,999	100.00	11,202	100.00
70 - Interview Complete	3,488	66.62	4,215	68.70	7,703	67.74
71 - No One at DU*	135	2.80	110	1.80	245	2.26
77 - Refusal	1,276	23.99	1,314	22.40	2,590	23.14
Other	304	6.58	360	7.10	664	6.86
Total						
Eligible Cases	42,942	100.00	44,714	100.00	87,656	100.00
70 - Interview Complete	32,869	71.24	35,440	74.71	68,309	73.04
71 - No One at DU*	1,366	3.60	1,226	2.47	2,592	3.01
77 - Refusal	6,004	19.86	5,484	17.48	11,488	18.63
Other	2,703	5.30	2,564	5.34	5,267	5.32

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.28 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Unweighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Hispanic or Latino								
Eligible Cases	4,936	100.00	4,954	100.00	4,016	100.00	13,906	100.00
70 - Interview Complete	4,294	86.99	3,945	79.63	2,929	72.93	11,168	80.31
71 - No One at DU*	82	1.66	228	4.60	223	5.55	533	3.83
77 - Refusal	125	2.53	619	12.49	693	17.26	1,437	10.33
Other	435	8.81	162	3.27	171	4.26	768	5.52
Not Hispanic or Latino Black or African American								
Eligible Cases	3,371	100.00	3,561	100.00	3,142	100.00	10,074	100.00
70 - Interview Complete	2,935	87.07	3,044	85.48	2,454	78.10	8,433	83.71
71 - No One at DU*	68	2.02	113	3.17	107	3.41	288	2.86
77 - Refusal	86	2.55	310	8.71	461	14.67	857	8.51
Other	282	8.37	94	2.64	120	3.82	496	4.92
Not Hispanic or Latino White								
Eligible Cases	16,436	100.00	17,650	100.00	22,288	100.00	56,374	100.00
70 - Interview Complete	13,316	81.02	13,829	78.35	16,020	71.88	43,165	76.57
71 - No One at DU*	271	1.65	658	3.73	624	2.80	1,553	2.75
77 - Refusal	619	3.77	2,722	15.42	5,019	22.52	8,360	14.83
Other	2,230	13.57	441	2.50	625	2.80	3,296	5.85
Not Hispanic or Latino Other or Multiple Races								
Eligible Cases	2,404	100.00	2,474	100.00	2,424	100.00	7,302	100.00
70 - Interview Complete	1,947	80.99	1,944	78.58	1,652	68.15	5,543	75.91
71 - No One at DU*	47	1.96	91	3.68	80	3.30	218	2.99
77 - Refusal	70	2.91	325	13.14	439	18.11	834	11.42
Other	340	14.14	114	4.61	253	10.44	707	9.68

Table 7.28 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Unweighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Large Metro								
Eligible Cases	12,060	100.00	12,868	100.00	14,168	100.00	39,096	100.00
70 - Interview Complete	9,949	82.50	9,998	77.70	9,971	70.38	29,918	76.52
71 - No One at DU*	225	1.87	550	4.27	537	3.79	1,312	3.36
77 - Refusal	423	3.51	1,920	14.92	3,024	21.34	5,367	13.73
Other	1,463	12.13	400	3.11	636	4.49	2,499	6.39
Small Metro								
Eligible Cases	9,264	100.00	10,218	100.00	10,768	100.00	30,250	100.00
70 - Interview Complete	7,704	83.16	8,232	80.56	7,923	73.58	23,859	78.87
71 - No One at DU*	142	1.53	354	3.46	300	2.79	796	2.63
77 - Refusal	298	3.22	1,346	13.17	2,214	20.56	3,858	12.75
Other	1,120	12.09	286	2.80	331	3.07	1,737	5.74
Nonmetro								
Eligible Cases	5,823	100.00	5,553	100.00	6,934	100.00	18,310	100.00
70 - Interview Complete	4,839	83.10	4,532	81.61	5,161	74.43	14,532	79.37
71 - No One at DU*	101	1.73	186	3.35	197	2.84	484	2.64
77 - Refusal	179	3.07	710	12.79	1,374	19.82	2,263	12.36
Other	704	12.09	125	2.25	202	2.91	1,031	5.63
Northeast								
Eligible Cases	5,513	100.00	6,114	100.00	6,674	100.00	18,301	100.00
70 - Interview Complete	4,421	80.19	4,720	77.20	4,632	69.40	13,773	75.26
71 - No One at DU*	100	1.81	289	4.73	257	3.85	646	3.53
77 - Refusal	192	3.48	943	15.42	1,520	22.77	2,655	14.51
Other	800	14.51	162	2.65	265	3.97	1,227	6.70

Table 7.28 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Unweighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Midwest								
Eligible Cases	7,733	100.00	7,891	100.00	8,875	100.00	24,499	100.00
70 - Interview Complete	6,399	82.75	6,270	79.46	6,473	72.94	19,142	78.13
71 - No One at DU*	131	1.69	303	3.84	263	2.96	697	2.85
77 - Refusal	272	3.52	1,092	13.84	1,866	21.03	3,230	13.18
Other	931	12.04	226	2.86	273	3.08	1,430	5.84
South								
Eligible Cases	8,292	100.00	8,583	100.00	9,404	100.00	26,279	100.00
70 - Interview Complete	6,973	84.09	7,012	81.70	6,901	73.38	20,886	79.48
71 - No One at DU*	154	1.86	318	3.70	324	3.45	796	3.03
77 - Refusal	214	2.58	1,010	11.77	1,810	19.25	3,034	11.55
Other	951	11.47	243	2.83	369	3.92	1,563	5.95
West								
Eligible Cases	5,609	100.00	6,051	100.00	6,917	100.00	18,577	100.00
70 - Interview Complete	4,699	83.78	4,760	78.66	5,049	72.99	14,508	78.10
71 - No One at DU*	83	1.48	180	2.97	190	2.75	453	2.44
77 - Refusal	222	3.96	931	15.39	1,416	20.47	2,569	13.83
Other	605	10.79	180	2.97	262	3.79	1,047	5.64
Male								
Eligible Cases	13,806	100.00	13,947	100.00	15,189	100.00	42,942	100.00
70 - Interview Complete	11,394	82.53	10,891	78.09	10,584	69.68	32,869	76.54
71 - No One at DU*	239	1.73	536	3.84	591	3.89	1,366	3.18
77 - Refusal	463	3.35	2,084	14.94	3,457	22.76	6,004	13.98
Other	1,710	12.39	436	3.13	557	3.67	2,703	6.29

Table 7.28 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Unweighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Female								
Eligible Cases	13,341	100.00	14,692	100.00	16,681	100.00	44,714	100.00
70 - Interview Complete	11,098	83.19	11,871	80.80	12,471	74.76	35,440	79.26
71 - No One at DU*	229	1.72	554	3.77	443	2.66	1,226	2.74
77 - Refusal	437	3.28	1,892	12.88	3,155	18.91	5,484	12.26
Other	1,577	11.82	375	2.55	612	3.67	2,564	5.73
Total								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.85	22,762	79.48	23,055	72.34	68,309	77.93
71 - No One at DU*	468	1.72	1,090	3.81	1,034	3.24	2,592	2.96
77 - Refusal	900	3.32	3,976	13.88	6,612	20.75	11,488	13.11
Other	3,287	12.11	811	2.83	1,169	3.67	5,267	6.01

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.29 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Weighted Percentages)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Hispanic or Latino								
Eligible Cases	4,936	100.00	4,954	100.00	4,016	100.00	13,906	100.00
70 - Interview Complete	4,294	86.97	3,945	79.67	2,929	71.38	11,168	74.95
71 - No One at DU*	82	1.47	228	4.44	223	5.36	533	4.67
77 - Refusal	125	3.01	619	12.57	693	18.46	1,437	15.33
Other	435	8.55	162	3.32	171	4.79	768	5.04
Not Hispanic or Latino Black or African American								
Eligible Cases	3,371	100.00	3,561	100.00	3,142	100.00	10,074	100.00
70 - Interview Complete	2,935	86.80	3,044	85.26	2,454	76.50	8,433	79.06
71 - No One at DU*	68	2.08	113	3.07	107	3.56	288	3.31
77 - Refusal	86	2.53	310	8.96	461	15.60	857	13.06
Other	282	8.59	94	2.71	120	4.34	496	4.57
Not Hispanic or Latino White								
Eligible Cases	16,436	100.00	17,650	100.00	22,288	100.00	56,374	100.00
70 - Interview Complete	13,316	80.69	13,829	78.12	16,020	70.47	43,165	72.19
71 - No One at DU*	271	1.65	658	3.49	624	2.56	1,553	2.59
77 - Refusal	619	3.65	2,722	15.78	5,019	23.26	8,360	20.79
Other	2,230	14.01	441	2.61	625	3.71	3,296	4.43
Not Hispanic or Latino Other or Multiple Races								
Eligible Cases	2,404	100.00	2,474	100.00	2,424	100.00	7,302	100.00
70 - Interview Complete	1,947	79.56	1,944	75.87	1,652	63.55	5,543	67.06
71 - No One at DU*	47	1.70	91	3.16	80	2.82	218	2.75
77 - Refusal	70	3.89	325	15.38	439	17.42	834	15.66
Other	340	14.85	114	5.60	253	16.21	707	14.53

Table 7.29 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Large Metro								
Eligible Cases	12,060	100.00	12,868	100.00	14,168	100.00	39,096	100.00
70 - Interview Complete	9,949	82.35	9,998	77.42	9,971	68.72	29,918	71.21
71 - No One at DU*	225	1.79	550	3.95	537	3.37	1,312	3.30
77 - Refusal	423	3.70	1,920	15.37	3,024	22.12	5,367	19.44
Other	1,463	12.16	400	3.26	636	5.78	2,499	6.06
Small Metro								
Eligible Cases	9,264	100.00	10,218	100.00	10,768	100.00	30,250	100.00
70 - Interview Complete	7,704	83.21	8,232	81.20	7,923	73.16	23,859	75.23
71 - No One at DU*	142	1.52	354	3.20	300	2.64	796	2.61
77 - Refusal	298	3.17	1,346	12.66	2,214	20.51	3,858	17.76
Other	1,120	12.10	286	2.94	331	3.70	1,737	4.40
Nonmetro								
Eligible Cases	5,823	100.00	5,553	100.00	6,934	100.00	18,310	100.00
70 - Interview Complete	4,839	83.79	4,532	82.08	5,161	72.99	14,532	75.05
71 - No One at DU*	101	1.59	186	3.11	197	2.95	484	2.84
77 - Refusal	179	2.62	710	12.48	1,374	20.04	2,263	17.53
Other	704	12.00	125	2.33	202	4.02	1,031	4.58
Northeast								
Eligible Cases	5,513	100.00	6,114	100.00	6,674	100.00	18,301	100.00
70 - Interview Complete	4,421	79.81	4,720	76.54	4,632	67.26	13,773	69.59
71 - No One at DU*	100	1.57	289	4.91	257	3.68	646	3.66
77 - Refusal	192	3.42	943	15.70	1,520	23.38	2,655	20.61
Other	800	15.19	162	2.85	265	5.67	1,227	6.15

Table 7.29 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Midwest								
Eligible Cases	7,733	100.00	7,891	100.00	8,875	100.00	24,499	100.00
70 - Interview Complete	6,399	83.34	6,270	79.64	6,473	72.22	19,142	74.27
71 - No One at DU*	131	1.62	303	3.57	263	2.87	697	2.84
77 - Refusal	272	3.11	1,092	13.97	1,866	21.12	3,230	18.44
Other	931	11.93	226	2.82	273	3.79	1,430	4.46
South								
Eligible Cases	8,292	100.00	8,583	100.00	9,404	100.00	26,279	100.00
70 - Interview Complete	6,973	83.52	7,012	81.70	6,901	71.75	20,886	74.22
71 - No One at DU*	154	2.03	318	3.70	324	3.31	796	3.23
77 - Refusal	214	2.55	1,010	11.60	1,810	20.18	3,034	17.32
Other	951	11.89	243	3.01	369	4.77	1,563	5.23
West								
Eligible Cases	5,609	100.00	6,051	100.00	6,917	100.00	18,577	100.00
70 - Interview Complete	4,699	83.44	4,760	77.22	5,049	70.61	14,508	72.75
71 - No One at DU*	83	1.24	180	2.50	190	2.45	453	2.34
77 - Refusal	222	4.85	931	16.90	1,416	21.55	2,569	19.30
Other	605	10.46	180	3.38	262	5.38	1,047	5.60
Male								
Eligible Cases	13,806	100.00	13,947	100.00	15,189	100.00	42,942	100.00
70 - Interview Complete	11,394	82.30	10,891	78.03	10,584	68.54	32,869	71.24
71 - No One at DU*	239	1.71	536	3.52	591	3.86	1,366	3.60
77 - Refusal	463	3.42	2,084	15.13	3,457	22.90	6,004	19.86
Other	1,710	12.57	436	3.32	557	4.69	2,703	5.30

Table 7.29 2012 Interview Results, by Age and Race/Ethnicity, Type of County, Region, and Gender (Weighted Percentages) (continued)

	12-17		18-25		26+		Total	
	Count	%	Count	%	Count	%	Count	%
Female								
Eligible Cases	13,341	100.00	14,692	100.00	16,681	100.00	44,714	100.00
70 - Interview Complete	11,098	83.39	11,871	80.49	12,471	72.75	35,440	74.71
71 - No One at DU*	229	1.64	554	3.68	443	2.38	1,226	2.47
77 - Refusal	437	3.32	1,892	13.10	3,155	19.85	5,484	17.48
Other	1,577	11.65	375	2.74	612	5.03	2,564	5.34
Total								
Eligible Cases	27,147	100.00	28,639	100.00	31,870	100.00	87,656	100.00
70 - Interview Complete	22,492	82.84	22,762	79.26	23,055	70.76	68,309	73.04
71 - No One at DU*	468	1.68	1,090	3.60	1,034	3.08	2,592	3.01
77 - Refusal	900	3.37	3,976	14.11	6,612	21.30	11,488	18.63
Other	3,287	12.12	811	3.03	1,169	4.87	5,267	5.32

DU = dwelling unit.

*Results include interviewer codes for no one at home after repeated visits and codes for respondent unavailable after repeated visits.

Table 7.30 2012 Interview Results—Spanish Interviews, by State (Unweighted Percentages)

	Spanish Interviews		English Interviews		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**
Total	2,061	3.02	66,248	96.98	68,309	100.00
Alabama	6	0.67	895	99.33	901	100.00
Alaska	0	0.00	829	100.00	829	100.00
Arizona	78	8.46	844	91.54	922	100.00
Arkansas	19	2.08	894	97.92	913	100.00
California	356	9.87	3,252	90.13	3,608	100.00
Colorado	36	3.88	891	96.12	927	100.00
Connecticut	21	2.18	943	97.82	964	100.00
Delaware	29	3.25	864	96.75	893	100.00
District of Columbia	33	3.43	929	96.57	962	100.00
Florida	271	7.65	3,273	92.35	3,544	100.00
Georgia	9	1.02	876	98.98	885	100.00
Hawaii	0	0.00	938	100.00	938	100.00
Idaho	13	1.41	908	98.59	921	100.00
Illinois	168	4.58	3,504	95.42	3,672	100.00
Indiana	9	0.99	902	99.01	911	100.00
Iowa	5	0.56	895	99.44	900	100.00
Kansas	11	1.21	901	98.79	912	100.00
Kentucky	2	0.22	925	99.78	927	100.00
Louisiana	8	0.89	893	99.11	901	100.00
Maine	0	0.00	938	100.00	938	100.00
Maryland	8	0.92	866	99.08	874	100.00
Massachusetts	50	5.24	905	94.76	955	100.00
Michigan	39	1.07	3,616	98.93	3,655	100.00
Minnesota	5	0.55	897	99.45	902	100.00
Mississippi	8	0.89	893	99.11	901	100.00

Table 7.30 2012 Interview Results—Spanish Interviews, by State (Unweighted Percentages) (continued)

	Spanish Interviews		English Interviews		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**
Missouri	5	0.55	910	99.45	915	100.00
Montana	0	0.00	876	100.00	876	100.00
Nebraska	18	1.91	922	98.09	940	100.00
Nevada	78	8.64	825	91.36	903	100.00
New Hampshire	3	0.32	947	99.68	950	100.00
New Jersey	59	6.57	839	93.43	898	100.00
New Mexico	57	6.48	822	93.52	879	100.00
New York	167	4.54	3,513	95.46	3,680	100.00
North Carolina	11	1.20	906	98.80	917	100.00
North Dakota	0	0.00	895	100.00	895	100.00
Ohio	10	0.27	3,677	99.73	3,687	100.00
Oklahoma	29	3.19	879	96.81	908	100.00
Oregon	20	2.17	903	97.83	923	100.00
Pennsylvania	34	0.95	3,546	99.05	3,580	100.00
Rhode Island	27	2.93	896	97.07	923	100.00
South Carolina	18	1.92	920	98.08	938	100.00
South Dakota	0	0.00	878	100.00	878	100.00
Tennessee	1	0.11	926	99.89	927	100.00
Texas	258	7.12	3,367	92.88	3,625	100.00
Utah	33	3.56	893	96.44	926	100.00
Vermont	3	0.34	882	99.66	885	100.00
Virginia	10	1.12	884	98.88	894	100.00
Washington	21	2.26	907	97.74	928	100.00
West Virginia	0	0.00	976	100.00	976	100.00
Wisconsin	12	1.37	863	98.63	875	100.00
Wyoming	3	0.32	925	99.68	928	100.00

Table 7.31 2012 Interview Results—Spanish Interviews, by State (Weighted Percentages)

	Spanish Interviews		English Interviews		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**
Total	2,061	4.55	66,248	95.45	68,309	100.00
Alabama	6	0.75	895	99.25	901	100.00
Alaska	0	0.00	829	100.00	829	100.00
Arizona	78	8.76	844	91.24	922	100.00
Arkansas	19	1.98	894	98.02	913	100.00
California	356	11.75	3,252	88.25	3,608	100.00
Colorado	36	5.00	891	95.00	927	100.00
Connecticut	21	1.92	943	98.08	964	100.00
Delaware	29	2.34	864	97.66	893	100.00
District of Columbia	33	3.09	929	96.91	962	100.00
Florida	271	10.48	3,273	89.52	3,544	100.00
Georgia	9	1.13	876	98.87	885	100.00
Hawaii	0	0.00	938	100.00	938	100.00
Idaho	13	1.36	908	98.64	921	100.00
Illinois	168	5.13	3,504	94.87	3,672	100.00
Indiana	9	1.25	902	98.75	911	100.00
Iowa	5	0.76	895	99.24	900	100.00
Kansas	11	1.36	901	98.64	912	100.00
Kentucky	2	0.14	925	99.86	927	100.00
Louisiana	8	1.13	893	98.87	901	100.00
Maine	0	0.00	938	100.00	938	100.00
Maryland	8	1.42	866	98.58	874	100.00
Massachusetts	50	3.74	905	96.26	955	100.00
Michigan	39	1.04	3,616	98.96	3,655	100.00
Minnesota	5	0.42	897	99.58	902	100.00
Mississippi	8	0.79	893	99.21	901	100.00

Table 7.31 2012 Interview Results—Spanish Interviews, by State (Weighted Percentages) (continued)

	Spanish Interviews		English Interviews		Total	
State	**Count**	**%**	**Count**	**%**	**Count**	**%**
Missouri	5	0.46	910	99.54	915	100.00
Montana	0	0.00	876	100.00	876	100.00
Nebraska	18	2.25	922	97.75	940	100.00
Nevada	78	9.01	825	90.99	903	100.00
New Hampshire	3	0.31	947	99.69	950	100.00
New Jersey	59	4.93	839	95.07	898	100.00
New Mexico	57	7.31	822	92.69	879	100.00
New York	167	5.03	3,513	94.97	3,680	100.00
North Carolina	11	1.07	906	98.93	917	100.00
North Dakota	0	0.00	895	100.00	895	100.00
Ohio	10	0.23	3,677	99.77	3,687	100.00
Oklahoma	29	3.33	879	96.67	908	100.00
Oregon	20	2.69	903	97.31	923	100.00
Pennsylvania	34	0.84	3,546	99.16	3,580	100.00
Rhode Island	27	2.82	896	97.18	923	100.00
South Carolina	18	1.43	920	98.57	938	100.00
South Dakota	0	0.00	878	100.00	878	100.00
Tennessee	1	0.39	926	99.61	927	100.00
Texas	258	9.81	3,367	90.19	3,625	100.00
Utah	33	2.80	893	97.20	926	100.00
Vermont	3	0.23	882	99.77	885	100.00
Virginia	10	1.23	884	98.77	894	100.00
Washington	21	3.77	907	96.23	928	100.00
West Virginia	0	0.00	976	100.00	976	100.00
Wisconsin	12	1.29	863	98.71	875	100.00
Wyoming	3	0.37	925	99.63	928	100.00

Table 7.32 2012 Interview Results—Spanish Interviews, by Age and Type of County (Unweighted Percentages)

	Spanish Interviews		English Interviews		Total	
	Count	%	Count	%	Count	%
Age Group						
12-17	441	1.96	22,051	98.04	22,492	100.00
18-25	544	2.39	22,218	97.61	22,762	100.00
26+	1,076	4.67	21,979	95.33	23,055	100.00
Type of County						
Large Metro	1,454	4.86	28,464	95.14	29,918	100.00
Small Metro	470	1.97	23,389	98.03	23,859	100.00
Nonmetro	137	0.94	14,395	99.06	14,532	100.00
Total	2,061	3.02	66,248	96.98	68,309	100.00

Table 7.33 2012 Interview Results—Spanish Interviews, by Age and Type of County (Weighted Percentages)

	Spanish Interviews		English Interviews		Total	
	Count	%	Count	%	Count	%
Age Group						
12-17	441	2.33	22,051	97.67	22,492	100.00
18-25	544	2.77	22,218	97.23	22,762	100.00
26+	1,076	5.21	21,979	94.79	23,055	100.00
Type of County						
Large Metro	1,454	6.52	28,464	93.48	29,918	100.00
Small Metro	470	3.00	23,389	97.00	23,859	100.00
Nonmetro	137	1.24	14,395	98.76	14,532	100.00
Total	2,061	4.55	66,248	95.45	68,309	100.00

Table 7.34 2012 English and Spanish Interviews Conducted, by Region

	Northeast		Midwest		South		West		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%
English	13,409	97.4	18,860	98.5	20,166	96.6	13,813	95.2	66,248	97.0
Spanish	364	2.6	282	1.5	720	3.4	695	4.8	2,061	3.0
Total	13,773	100.0	19,142	100.0	20,886	100.0	14,508	100.0	68,309	100.0

Table 7.35 2012 English and Spanish Interviews Conducted, by Population Density

	1,000,000		50,000-999,999		Non-CBSA		Total	
	Count	%	Count	%	Count	%	Count	%
English	27,005	95.0	33,921	98.2	5,322	99.4	66,248	97.0
Spanish	1,410	5.0	617	1.8	34	0.6	2,061	3.0
Total	28,415	100.0	34,538	100.0	5,356	100.0	68,309	100.0

CBSA = core-based statistical area.

Table 7.36 2012 Interviewer's Assessment of Respondent's Level of Understanding, by Age and Race/Ethnicity of Respondent

Interviewer Assessment	12-17	18-25	26+	Total
Hispanic or Latino				
Total Number	4,442	3,989	3,052	11,483
Level of Understanding (Percent of Total)				
No Difficulty	92.5	92.7	81.9	89.8
Just a Little Difficulty	6.2	5.9	13.2	7.9
A Fair Amount of Difficulty	1.1	1.1	3.8	1.8
A Lot of Difficulty	0.3	0.3	1.0	0.5
No Response	0.0	0.0	0.1	0.0
Not Hispanic or Latino Black or African American				
Total Number	2,916	2,998	2,507	8,421
Level of Understanding (Percent of Total)				
No Difficulty	92.6	95.0	88.2	92.2
Just a Little Difficulty	5.6	3.7	8.1	5.7
A Fair Amount of Difficulty	1.2	0.9	2.8	1.6
A Lot of Difficulty	0.5	0.4	1.0	0.6
No Response	0.0	0.0	0.0	0.0
Not Hispanic or Latino White				
Total Number	13,047	13,523	15,983	42,553
Level of Understanding (Percent of Total)				
No Difficulty	94.8	96.8	93.5	94.9
Just a Little Difficulty	4.2	2.6	5.0	4.0
A Fair Amount of Difficulty	0.8	0.4	1.1	0.8
A Lot of Difficulty	0.2	0.2	0.4	0.3
No Response	0.0	0.1	0.0	0.0
Not Hispanic or Latino Other or Multiple Races				
Total Number	2,068	2,019	1,765	5,852
Level of Understanding (Percent of Total)				
No Difficulty	92.4	93.1	85.4	90.5
Just a Little Difficulty	6.2	5.4	10.8	7.3
A Fair Amount of Difficulty	0.9	1.1	2.9	1.6
A Lot of Difficulty	0.4	0.4	0.8	0.5
No Response	0.1	0.0	0.0	0.1

Table 7.37 2012 Interviewer's Assessment of Respondent's Level of Cooperation during Interview, by Age and Race/Ethnicity of Respondent

Interviewer Assessment	12-17	18-25	26+	Total
Hispanic or Latino				
Total Number	4,442	3,989	3,052	11,483
Level of Cooperation (Percent of Total)				
Very Cooperative	98.1	97.3	95.7	97.2
Fairly Cooperative	1.7	2.4	3.9	2.5
Not Very Cooperative	0.1	0.2	0.2	0.2
Openly Hostile	0.0	0.0	0.2	0.1
No Response	0.0	0.0	0.1	0.0
Not Hispanic or Latino Black or African American				
Total Number	2,916	2,998	2,507	8,421
Level of Cooperation (Percent of Total)				
Very Cooperative	98.5	97.8	96.9	97.8
Fairly Cooperative	1.4	2.0	2.8	2.0
Not Very Cooperative	0.1	0.2	0.2	0.2
Openly Hostile	0.0	0.0	0.0	0.0
No Response	0.0	0.0	0.0	0.0
Not Hispanic or Latino White				
Total Number	13,047	13,523	15,983	42,553
Level of Cooperation (Percent of Total)				
Very Cooperative	98.8	98.6	98.0	98.4
Fairly Cooperative	1.1	1.2	1.8	1.4
Not Very Cooperative	0.1	0.1	0.1	0.1
Openly Hostile	0.0	0.0	0.1	0.0
No Response	0.0	0.1	0.0	0.0
Not Hispanic or Latino Other or Multiple Races				
Total Number	2,068	2,019	1,765	5,852
Level of Cooperation (Percent of Total)				
Very Cooperative	98.2	98.2	97.2	97.9
Fairly Cooperative	1.6	1.5	2.6	1.9
Not Very Cooperative	0.1	0.1	0.2	0.1
Openly Hostile	0.0	0.1	0.1	0.1
No Response	0.1	0.0	0.0	0.1

Table 7.38 2012 Interviewer's Assessment of Respondent's Level of Privacy during Interview, by Age and Race/Ethnicity of Respondent

Interviewer Assessment	12-17	18-25	26+	Total
Hispanic or Latino				
Total Number	4,442	3,989	3,052	11,483
Level of Privacy (Percent of Total)				
Completely Private	73.4	80.9	80.0	77.8
Minor Distractions	19.5	14.5	15.5	16.7
Person(s) in Room 1/3 of Time	2.6	1.8	1.6	2.1
Serious Interruptions > 1/2 of Time	0.6	0.4	0.5	0.5
Constant Presence of Other Person(s)	3.9	2.4	2.3	3.0
Not Sure	0.0	0.0	0.1	0.0
Not Hispanic or Latino Black or African American				
Total Number	2,916	2,998	2,507	8,421
Level of Privacy (Percent of Total)				
Completely Private	76.4	84.0	85.1	81.7
Minor Distractions	16.9	11.9	10.2	13.1
Person(s) in Room 1/3 of Time	2.4	1.5	2.0	2.0
Serious Interruptions > 1/2 of Time	0.6	0.5	0.3	0.5
Constant Presence of Other Person(s)	3.6	2.1	2.4	2.7
Not Sure	0.0	0.0	0.0	0.0
Not Hispanic or Latino White				
Total Number	13,047	13,523	15,983	42,553
Level of Privacy (Percent of Total)				
Completely Private	77.2	84.9	86.3	83.1
Minor Distractions	16.9	11.3	10.3	12.6
Person(s) in Room 1/3 of Time	2.2	1.3	1.3	1.6
Serious Interruptions > 1/2 of Time	0.5	0.4	0.3	0.4
Constant Presence of Other Person(s)	3.2	2.1	1.7	2.3
Not Sure	0.0	0.1	0.0	0.0
Not Hispanic or Latino Other or Multiple Races				
Total Number	2,068	2,019	1,765	5,852
Level of Privacy (Percent of Total)				
Completely Private	73.5	82.6	78.7	78.2
Minor Distractions	19.7	13.5	16.1	16.5
Person(s) in Room 1/3 of Time	2.9	1.6	2.2	2.2
Serious Interruptions > 1/2 of Time	0.5	0.2	0.7	0.5
Constant Presence of Other Person(s)	3.2	2.0	2.3	2.5
Not Sure	0.1	0.0	0.0	0.1

Table 7.39 2012 Interviewer's Assessment of How Often Respondent Revealed Answers in ACASI Sections, by Age and Race/Ethnicity of Respondent

Interviewer Assessment	12-17	18-25	26+	Total
Hispanic or Latino				
Total Number	4,442	3,989	3,052	11,483
How Often Reveal Answer (Percent of Total)				
None of the Time	96.8	96.4	88.1	94.3
A Little of the Time	2.9	3.1	9.2	4.6
Some of the Time	0.1	0.2	1.8	0.6
A Lot of the Time	0.1	0.1	0.6	0.2
All of the Time	0.1	0.1	0.2	0.1
No Response	0.0	0.0	0.1	0.0
Not Hispanic or Latino Black or African American				
Total Number	2,916	2,998	2,507	8,421
How Often Reveal Answer (Percent of Total)				
None of the Time	96.9	98.1	91.3	95.7
A Little of the Time	2.6	1.8	6.2	3.4
Some of the Time	0.4	0.1	1.5	0.6
A Lot of the Time	0.0	0.0	0.6	0.2
All of the Time	0.1	0.1	0.3	0.2
No Response	0.0	0.0	0.0	0.0
Not Hispanic or Latino White				
Total Number	13,047	13,523	15,983	42,553
How Often Reveal Answer (Percent of Total)				
None of the Time	96.9	97.0	92.6	95.3
A Little of the Time	2.6	2.4	5.8	3.8
Some of the Time	0.1	0.2	0.9	0.4
A Lot of the Time	0.1	0.1	0.3	0.2
All of the Time	0.3	0.2	0.4	0.3
No Response	0.0	0.1	0.0	0.0
Not Hispanic or Latino Other or Multiple Races				
Total Number	2,068	2,019	1,765	5,852
How Often Reveal Answer (Percent of Total)				
None of the Time	97.0	96.3	90.1	94.7
A Little of the Time	2.4	3.2	7.9	4.3
Some of the Time	0.4	0.3	1.3	0.6
A Lot of the Time	0.0	0.0	0.5	0.2
All of the Time	0.1	0.1	0.2	0.1
No Response	0.1	0.0	0.0	0.1

ACASI = audio computer-assisted self-interviewing.

Table 7.40 Number of Visits Required To Complete Screening

Visits	Screenings	%	Cumulative %
1	67,384	31.4	31.4
2	43,160	20.1	51.6
3	25,655	12.0	63.6
4	17,268	8.1	71.6
5-9	38,102	17.8	89.4
10+	22,704	10.6	100.0
Missing	1	0.0	100.0
Total	214,274		

Table 7.41 Number of Visits Required To Complete Interview

Visits	Interviews	%	Cumulative %
1	24,588	36.0	36.0
2	23,380	34.2	70.2
3	7,898	11.6	81.8
4	3,675	5.4	87.2
5-9	6,299	9.2	96.4
10+	2,358	3.5	99.8
Missing	111	0.2	100.0
Total	68,309		

8. Quality Control

While every step was designed to help collect the highest quality data possible, the 2012 National Survey on Drug Use and Health (NSDUH) included specific quality control processes, which are described in this chapter.

8.1 Field Supervisor and Interviewer Evaluation

8.1.1 Regular Conferences

Each field interviewer (FI) had at least one regularly scheduled weekly telephone conference with his or her field supervisor (FS). During this call, the FI reported progress made toward completing the work; reviewed production, time, and expense information for the week; discussed field problems; and asked any questions that had emerged during the week. The FS provided feedback on the progress and quality of work and offered solutions to problems or questions encountered. The FS also shared any information from project managers, such as approaching project deadlines.

Regular weekly telephone conferences were also held between the regional supervisor (RS) and each of the FSs in his or her territory. FI production and performance were discussed during these conferences, as were budget considerations, cost containment issues, and any problems that were occurring.

8.1.2 New-to-Project Training and Training Evaluations

Beginning at new-to-project training, FI performance was monitored closely and consistently throughout the field period. Training classes were small enough to observe and evaluate each FI's individual performance and comprehension. The classroom trainers worked together to evaluate FIs on a daily basis using the Daily FI Training Evaluation (see Exhibit 5.1).

Through the certification process (see Section 5.2.1), a formal one-on-one evaluation of each FI occurred. As explained earlier, all FIs were required to pass an individually conducted certification in order to successfully complete training.

In addition, all new-to-project graduates were mentored (see Section 5.2.5) to observe their behavior in the field and reinforce the important study protocols learned during training.

8.1.3 Veteran Training and Ongoing FI Knowledge Evaluations

Veteran FIs continuing work on the study in 2012 were tested and trained to be sure they met the standards necessary to serve as NSDUH interviewers. Beginning with the iLearning courses (see Section 4.6.1), interviewers could only continue working if they demonstrated knowledge of basic protocols, successfully completed all veteran training iLearning courses, and attended their assigned FS team meeting and training session.

Periodic evaluations of interviewer knowledge occurred during the year as FIs completed the "Habits for Success" iLearning course prior to the start of each quarter (see Section 5.5). This tool not only tested knowledge but reinforced that following protocol helped collect data of the highest possible quality. All interviewers also received a Showcard Booklet containing the "Steps to Maximize Data Quality" (see Exhibit 8.1), which listed the most crucial NSDUH protocol steps.

8.1.4 Field Interviewer Observations

In-person observations of FIs at work provided insights about the survey and its procedures as well as assessments of interviewer performance and attention to project protocol. Field observations were conducted nationally in all four quarters of 2012.

Around the country, a total of 180 field observations were completed, in which 160 different FIs were observed completing 448 screenings and 265 interviews. Observers, who were RSs, FSs, regional directors (RD), instrumentation team members, project survey specialists, other RTI staff members, or Substance Abuse and Mental Health Services Administration (SAMHSA) staff, had specific forms to complete, noting interviewer behaviors on a number of project protocols.

To maintain the integrity of the operation, observers did not give direct feedback to the FIs. Information regarding FI performance was made available to the appropriate FS to share with observed FIs through structured feedback and retraining plans. Results from these observations were formally documented in the 2012 NSDUH Full-Year Field Observation Report.

8.1.5 FS Evaluations of FIs

To streamline year-end field management responsibilities and since appropriate and timely feedback occurred throughout the year, FSs were not required to complete an annual evaluation of their FIs. Instead, when an interviewer left the project, the FS was required to complete a Field Data Collector Evaluation Form, documenting the strengths and weaknesses of the interviewer. Completed evaluations were added to the interviewer's personal data file at RTI. The FS generally completed this form without RS or RD input.

8.1.6 FI Exit Interviews

Section 8.1.6 text has been removed.

8.2 Data Quality Team

The Data Quality Team was responsible for the identification, resolution, and distribution of information to field management staff concerning data quality and verification issues. Following a reorganization implemented in Quarter 2 of 2012 as part of the Data Quality Process Improvement initiative, a NSDUH methodologist served as the data quality director, reporting directly to the project director and providing oversight for the team of two data quality managers (DQMs) and a verification coordinator. The DQMs closely monitored the data quality of designated RS areas, identifying trends in data quality errors and indications of potential falsified work. The verification coordinator, also on the operations team, was responsible for operational tasks associated with the verification process such as overseeing the call center and telephone verification activities. To ensure reliable succession planning and backup, the verification coordinator was also trained on the DQM role and was responsible for DQM tasks associated with one RS area.

The Data Quality Team also developed the "Data Quality Knowledge Notebook," which summarized procedures for various topics. Each quarter, the Data Quality Team identified six topics for all FIs to review. FIs are retrained on the quarter's six data quality topics by reading the applicable section(s) of the FI Manual for each topic, reviewing a list of key points in the notebook, and completing notebook questions to demonstrate their data quality knowledge related to the topic. In addition, each FS reviewed the six topics with their FIs during either individual or team conference calls.

8.3 Data Quality Monitoring

The NSDUH web-based Case Management System (CMS) enabled the Data Quality Team and field management staff to monitor the quality of each FI's work through case reports and other functions generated from daily data transmissions from the FIs' iPAQs and laptops. Reports focusing on data quality problems generated from these data were provided by type of problem and FI.

Access to the data quality reports varied by the project responsibilities of each staff member.

8.3.1 Field Management Data Quality Reports

Reports were available for review and analysis by field supervisory staff, project management staff, and the Data Quality Team so corrective actions could be taken as necessary. The information contained in these reports was addressed during weekly conference calls between FSs and FIs and between RSs and FSs.

8.3.2 Data Quality Team Data Quality Reports

Field supervisory staff focused their efforts on the field management data quality reports , while reports providing additional details or requiring more expertise for proper analysis were available for the Data Quality Team.

All of the reports were carefully reviewed by DQMs to identify trends in data quality errors and indications of potential falsified work. DQMs highlighted these trends and discussed them with other field management staff members to determine if further investigation or increased verification of an FI's work, FI retraining, and/or disciplinary action against the FI should occur. In addition, these reports were used to identify recurring data quality problems that warranted further discussion with FIs via "Data Quality Knowledge Notebook" topics, iLearning refresher courses, and FI retraining sessions.

8.4 Verification of Completed Cases

In order to verify the quality and accuracy of each FI's work, a complex verification procedure was implemented. This involved the selection and verification of a percentage of final interview cases, as well as a percentage of final noninterview screening cases for each interviewer. Verification contacts for selected cases were made primarily by telephone.

The system allowed for the verification of additional work beyond the standard selection rates up to 100 percent of the FI's completed work.

8.4.1 In-House Verification

Contact information used in the verification process for completed interviews was obtained from the Quality Control Form completed by each interview respondent (see Exhibit 8.5). For the final noninterview screening codes of 10 (vacant), 13 (not primary residence), 18 (not a dwelling unit), 22 (dwelling unit contains only military personnel), 26 (not eligible for the quarter), and 30 (no one selected for interview), the contact information was recorded in the iPAQ at the time the case was finalized. For codes 10, 13, and 18, the contact was made with a knowledgeable person, such as a real estate agent, property manager, or neighbor. For codes 22, 26, and 30, the verification was completed most often with the screening respondent.

The telephone verification was conducted by project-trained telephone interviewers in RTI's Call Center Services (CCS) unit. Spanish translations of all materials were available for verifications with Spanish-speaking respondents.

The NSDUH telephone verification script used depended on the final status code of the case (see Appendix E).

Telephone interviewers followed a script when speaking with the respondent to confirm that the FI was professional and followed project protocols. The majority of cases were finalized as having no problems.

Tables 8.3 and 8.4 provide summaries of the results of phone verifications for noninterview screening codes 10, 13, 18, 22, 26, and 30 and for completed interviews.

8.4.2 Field Verification

In addition to the verification procedures conducted on completed work received in-house, additional steps were taken in the field to ensure complete and accurate collection of data.

The Data Quality Team worked with the RD as needed to select the cases to be field verified.

The Field Verifier returned to the SDUs that were assigned and queried the respondents in an effort to determine whether or not proper contact had been made by the FI in question.

The Field Verifier also reviewed some protocol issues with the respondent to ensure the FI had followed protocol and acted in a professional manner. Results of the field

verification were reported to the Data Quality Team and the FS, RS, RD, National Field Director, associate project officers, and project director. If the Field Verifier found the work to be invalid, he or she reworked the case.

In the 2012 NSDUH, a total of

260 interviews and 557 screenings were determined to have been falsified. All 22 FIs with falsified work were terminated. In addition to being terminated, any FI who falsified work is no longer eligible to work on NSDUH or any other study conducted by RTI or Headway. All of the invalid cases were either reworked or removed from the dataset, as were any unverifiable cases completed by FIs with falsified work.

In other 2012 field verification findings, 13 FIs made errors to cause a total of 33 screenings and 16 interviews to be invalid, while 11 other FIs made errors without causing their work to be invalid. In these situations, no clear evidence of falsification was found. Sixteen of these FIs were retrained and returned to work; five FIs were placed on probation and retrained and returned to work; and three FIs resigned.

No evidence of errors or falsification was found for the 18 remaining FIs, and they could return to work without retraining. Sixteen of these FIs returned to work and two FIs resigned. Tables 8.5, 8.6, and 8.7 provide summaries of the field verification results for selected screening cases, interview cases, and FIs.

8.4.3 Verification Reporting Tools

8.4.3.1 Case Data Information Link

Project staff could view the Verification Status of each case through the Case Data Information link on the CMS.

8.4.3.2 Short FI Level Verification Report

The Short FI Level Verification Report provided a snapshot of the problems identified during Verification to the Data Quality Team and other key field management staff. The main table provided a summary of verification data.

On page 2 of the report, more specific details of the problems identified during Verification were displayed in tables based on the result code of the case.

8.4.3.3 Field Verification Summary Report

The Field Verification Summary Report provided a summary of problems found during field verification to project staff. The number of cases selected for field verification was displayed along with the results.

8.5 Industry and Occupation Coding

During the latter part of the interview, the FI asked a series of questions to obtain detailed information about a respondent's job. Quarterly, RTI sent this information to The National Processing Center of the U.S. Census Bureau so that their team of industry and occupation coders could classify each respondent's job. Details on the end results from the U.S. Census Bureau coding operation are provided in Appendix F.

All interviewers had available in the Showcard Booklet a listing of tips and helpful hints to use when collecting industry and occupation data. Based on prior experience, common problem situations were included to provide examples of the level of detail required to assign codes.

Table 8.1 2012 NSDUH FI Exit Interview Results

Table has been removed.

Table 8.1 2012 NSDUH FI Exit Interview Results (continued)

Table has been removed.

Table 8.1 2012 NSDUH FI Exit Interview Results (continued)

Table has been removed.

Table 8.1 2012 NSDUH FI Exit Interview Results (continued)

Table has been removed.

Table 8.1 2012 NSDUH FI Exit Interview Results (continued)

Table has been removed.

Table 8.1 2012 NSDUH FI Exit Interview Results (continued)

Table has been removed.

Table 8.2 2012 NSDUH FI Exit Interviews—Most Important Reason for Resignation

Table has been removed.

Table 8.3 2012 NSDUH Phone Verification Results—Noninterview Screening Cases

Table has been removed.

Table 8.4 2012 NSDUH Phone Verification Results—Interview Cases

Table has been removed.

Table 8.5 2012 NSDUH Field Verification Results—Noninterview Screening Cases

Table has been removed.

Table 8.6 2012 NSDUH Field Verification Results—Interview Cases

Table has been removed.

Table 8.7 2012 NSDUH Field Verification Results—Field Interviewers

Table has been removed.

Exhibit 8.1 Steps to Maximize Data Quality

Steps to Maximize Data Quality

This summary is not a replacement for information contained in your FI Manual, but is a listing of some of our most crucial protocols that must be followed.

Note the FI Manual pages referenced with each key point. Keep in mind these protocols are not the only steps necessary to follow. Use your FI Manual, Field Supervisor, and project e-mails for information on additional steps to maximize data quality.

BE SURE YOU FOLLOW EACH OF THESE PROTOCOLS AT ALL TIMES.

SCREENING

- **Use your segment maps**, and not just the address, to locate your selected DUs. [FI Manual p. 3-21]
- **Display your ID badge** when knocking on every door in your segment. [FI Manual pgs. 4-15 and 5-1]
- **Complete screenings in-person** with a resident who is 18 or older. The only exception is in the case of emancipated minors. [FI Manual p. 4-16]
- **Give a Study Description to each SR.** [FI Manual p. 4-17]
- Obtain complete and accurate screening information, **reading the screening questions verbatim to the SR** and immediately entering responses into the iPAQ. The only missing screening data should be a result of the respondent's refusal or inability to provide the information. [FI Manual p. 6-15]

INTERVIEW

- **Read the CAI Introduction and Informed Consent from the Showcard Booklet** to the R (choosing the appropriate version based on the respondent's age) before beginning the interview. Before speaking with a selected minor, you must obtain verbal parental permission. **If the R was not the SR, give him/her a Study Description.** [FI Manual pgs. 7-22 and 7-23]
- Make it apparent you are **completing the interview in a completely confidential and unbiased manner**. [FI Manual pgs. 2-6, 2-7 and 8-1]
- To the extent possible, **choose an interview location that gives the respondent privacy**. [FI Manual pgs. 7-26 and 7-27]

Exhibit 8.1 Steps to Maximize Data Quality (continued)

INTERVIEW, CONT.

- **Do not rush the respondent**. Do not tell the respondent how to make the interview go faster. [FI Manual p. 8-2 and 8-3]
- Use the **Reference Date Calendar and read the explanation provided on the CAI screens verbatim to the R**. As appropriate, remind the respondent to use the calendar as a visual aid throughout the interview. [FI Manual p. 8-14]
- Familiarize the R with the laptop and function keys by reading the provided script in the CAI Interview and **allow the R to successfully complete the Computer Practice on his or her own**. You must always explain, offer, AND plug in the headphones with each R. [FI Manual pgs. 8-17 through 8-20]
- **Read the interview questions exactly as they appear on the screen**. It is never acceptable to use your own words or 'wing it'. Do not assume you know answers from a previous conversation, question, or interview. [FI Manual pgs. 8-2 and 8-3]
- **Hand the appropriate Showcard to the respondent** when instructed on the CAI screen. [FI Manual p. 8-13]
- Allow your respondents to complete the ACASI portion of the interview on their own. **Never read the questions in the ACASI portion out loud to the respondent**. In cases of extreme physical impairment, it may be necessary to enter the answers into the computer for the ACASI questions, but always allow the ACASI recording to 'read' the questions and answer categories via the headphones. [FI Manual p. 8-24]
- **Have the respondent fill out the top portion of the Quality Control Form** and allow him or her to insert the form into the envelope and seal it. Mail the form within 24 hours of the interview. [FI Manual pgs. 8-26 through 8-27]
- Always protect the confidentiality of your respondents. **Never reveal a respondent's answers to anyone**, including the respondent's family members. Resist the temptation to reveal even positive information gleaned from an interview to parents or other household members. [FI Manual pgs. 2-6 through 2-8]

Exhibit 8.2 2012 NSDUH Field Interviewer Exit Interview

Exhibit has been removed.

Exhibit 8.2 2012 NSDUH Field Interviewer Exit Interview (continued)

Exhibit has been removed.

Exhibit 8.2 2012 NSDUH Field Interviewer Exit Interview (continued)

Exhibit has been removed.

Exhibit 8.2 2012 NSDUH Field Interviewer Exit Interview (continued)

Exhibit has been removed.

Exhibit 8.2 2012 NSDUH Field Interviewer Exit Interview (continued)

Exhibit has been removed.

Exhibit 8.3 Overview of NSDUH Noninterview Screening Verification Process

Exhibit has been removed.

Exhibit 8.4 Overview of NSDUH Interview Verification Process

Exhibit has been removed.

Exhibit 8.5 Quality Control Form

VERSIÓN EN ESPAÑOL AL REVERSO

NOTICE: Public reporting burden (or time) for this collection of information is estimated to average 2 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to SAMHSA Reports Clearance Officer, Paperwork Reduction Project (0930-0110); Room 8-1099; 1 Choke Cherry Road, Rockville, MD 20857. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this project is 0930-0110.

OMB No.: 0930-0110
OMB Expiration Date: 08-31-14

QUALITY CONTROL FORM

As part of our quality control program, we plan to contact a portion of the survey participants to make sure that the interviewer has followed the study procedures. We only ask general questions—no specific information is required. We sincerely appreciate your cooperation.

Please fill in the boxes below. (PLEASE PRINT CLEARLY.) **Thank you.**

[Your phone number will be kept confidential and will not be released to anyone other than our quality control representatives.]

TELEPHONE NUMBER				–				–				
	(Area Code)				**(Telephone Number)**							

YOUR ADDRESS	

CITY		**STATE**			**ZIP CODE**					

BOXES BELOW MUST FIRST BE COMPLETED [IN INK] BY INTERVIEWER.																
TODAY'S DATE	M	M	-	D	D	-	1	2	**TIME**			:			AM PM	
FI NAME									**FI ID #**							
CASE ID #							–		–				–		**Include A or B!**	
IF respondent is 12 – 17 years old, which adult granted permission for the interview? (Examples: father, mother, etc.)	[Print Parent/Guardian's relationship to the child in this box.]															

Exhibit 8.5 Quality Control Form (continued)

ENGLISH VERSION ON OTHER SIDE

NOTA: Se calcula que el tiempo que le tomará a cada participante para dar esta información será 2 minutos, incluyendo el tiempo para repasar las instrucciones, buscar las fuentes de información existentes, reunir y mantener los datos requeridos, así como completar y revisar la recopilación de información. Envíe sus comentarios acerca de este cálculo de tiempo o cualquier otro aspecto relacionado con esta recolección de información, incluyendo sugerencias para reducir el tiempo a: SAMHSA Reports Clearance Officer, Paperwork Reduction Project (0930-0110); Room 8-1099; 1 Choke Cherry Road, Rockville, MD 20857. Ninguna agencia está autorizada a realizar o patrocinar ninguna recopilación de información sin presentar un número de control válido de la Oficina de Administración y Presupuesto (OMB, por sus siglas en inglés), ni tampoco está obligada ninguna persona a participar en una recopilación de información si no existe dicho número. El número de control OMB para este proyecto es 0930-0110.

No. de control OMB: 0930-0110
Fecha de vencimiento: 31 de agosto de 2014

FORMULARIO DE CONTROL DE CALIDAD

Como parte de nuestro programa de control de calidad, pensamos comunicarnos con un grupo de participantes de esta encuesta para asegurarnos que el (la) entrevistador(a) ha cumplido con los procedimientos apropiados del estudio. Sólo haremos preguntas en general y no solicitaremos ninguna información específica. Le agradecemos sinceramente su colaboración.

Por favor llene los espacios en blanco a continuación. (FAVOR DE ESCRIBIR CLARAMENTE.) **Gracias.**

[Su número de teléfono se mantendrá confidencial y sólo se dará esta información a nuestro personal encargado del control de calidad.]

NÚMERO DE TELÉFONO				–				–				
	(Código de área)				**(Número de teléfono)**							

SU DOMICILIO	

CIUDAD		**ESTADO**			**CÓDIGO POSTAL**					

BOXES BELOW MUST FIRST BE COMPLETED [IN INK] BY INTERVIEWER.																	
TODAY'S DATE	M	M	-	D	D	–	1	2	**TIME**			:			AM PM		
FI NAME									**FI ID #**								
CASE ID #							–		–				–		**Include A or B!**		
IF respondent is 12 – 17 years old, which adult granted permission for the interview? (Examples: father, mother, etc.)	[Print Parent/Guardian's relationship to the child in this box.]																

Exhibit 8.6 CAI Mail Verification Letter

Exhibit has been removed.

Exhibit 8.7 Short FI-Level Verification Report—Page 1

Exhibit has been removed.

Exhibit 8.7 Short FI-Level Verification Report—Page 1 (continued)

Exhibit has been removed.

Exhibit 8.8 Short FI-Level Verification Report—Page 2

Exhibit has been removed.

Exhibit 8.8 Short FI-Level Verification Report—Page 2 (continued)

Exhibit has been removed.

Exhibit 8.9 Short FI-Level Verification Report Problem Codes

Exhibit has been removed.

Exhibit 8.9 Short FI-Level Verification Report Problem Codes (continued)

Exhibit has been removed.

Exhibit 8.10 Field Verification Summary Report

Exhibit has been removed.

Reference

Clark, C. M., Cohen, L. M., Etzel, K. C., Johnson, I., Parsons, S. L., Rumsey, A. M., & Touarti, C. M. (2011, August). *2012 National Survey on Drug Use and Health—Data collection quality control plan* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 10). Research Triangle Park, NC: RTI International.

Appendix A: New-to-Project Home Study Cover Memo

RTI INTERNATIONAL | Memorandum

TO: NSDUH New-to-Project Field Interviewers

FROM: Ilona Johnson, National Field Director

RE: 2012 NSDUH Home Study Package

DATE: December 27, 2011

Thank you for your interest in the 2012 National Survey on Drug Use and Health (NSDUH). We are excited to have you join this important research study. Please carefully follow all the instructions provided for completing the New-to-Project (NTP) eHome Study and preparing for the NSDUH Field Interviewer (FI) training session. **You must complete the NTP eHome Study by 11:59 PM Eastern Standard Time on Thursday, January 12, 2012 and score at least 80%** to attend this training session.

I. PREPARATIONS FOR THE NTP eHOME STUDY

In addition to this memo, your shipment includes the materials listed below to prepare you for the upcoming training session. If you are missing any items, please let your Field Supervisor (FS) know right away.

- **2012 NSDUH FI Manual:** a 3-ring binder outlining specific protocols and procedures you must follow to complete your NSDUH assignment.
- **2012 NSDUH FI Computer Manual:** outlines protocols and procedures for the use and care of your NSDUH computer equipment. *(Your computer equipment will be issued at training*.)
- **NTP eHome Study (paper version):** use this for reference as you review your manuals and as a guide when you complete the NTP eHome Study via the internet.
- **Background Investigation Requirements memo:** provides additional information on the background investigative requirements for FIs hired on NSDUH.
- **2012 Mental Health Surveillance Study FI Handbook:** The Mental Health Surveillance Study is a special research study added to the 2012 NSDUH. This handbook outlines all procedures and materials specific to this study and should be reviewed prior to training.

II. COMPLETING THE NTP eHOME STUDY VIA THE INTERNET

- You may complete the NTP eHome Study on any computer with internet access, whether it's in your home, a friend's house, the public library, etc. You will only need basic computer skills, such as "pointing and clicking" the mouse and occasionally scrolling down the page. All other instructions are included on the screen within the eHome Study. This is an un-timed, open-book exercise, so take your time and refer to the manuals when answering questions, as necessary.
- Your FS will provide you with your FI ID number that you will need to access the eHome Study. After conducting a thorough review of the manuals, it should take no more than 45 minutes to complete the entire eHome Study. After you submit your eHome Study online, your FS will receive your score and contact you within a few days to let you know how you did.
- **In order to attend training, you must achieve a passing score of least 80% on the eHome Study** (or answer 35 out of 44 questions correctly). Anyone who misses 10 questions or more will fail the eHome Study and will not be allowed to attend training.
- Please submit your completed eHome Study via the internet by **11:59 PM Eastern Standard Time on Thursday, January 12, 2012**. If you miss this deadline, you cannot attend training.
- The eHome Study consists of 48 questions and is divided into three sections: Section 1 - FI Manual (questions 1-34); Section 2 - FI Computer Manual (questions 35-44); and Section 3 - General Internet (questions 45-48 – *not graded*).

ENTERING YOUR ANSWERS & SUBMITTING VIA THE INTERNET

- Access the internet by opening an internet browser (such as Microsoft Internet Explorer). If needed, feel free to ask a relative or friend to help you.
- Start by going to this website: **https://nsduhweb.rti.org/homestudy/login.cfm**
  To do this, type the bolded text above in the "Address" window of your internet browser. This will take you to the entry screen, shown below.

- On the entry screen type your FI ID, given to you by your FS, in the box provided. Then click the grey button labeled "Login."
- This will take you to the screen shown below to confirm your name. If the information is correct and you see your name displayed, click "Yes." If the information is incorrect, click "Cancel" to re-enter your FI ID.

- After clicking "Yes," you will have successfully opened the NSDUH eHome Study and can begin answering the questions. To enter your responses, click the white circle next to the best answer category. Only one response can be given for each question. Once you have completed all the questions on a screen, click "Next" to advance to the next screen and a new set of questions. Continue this process until the eHome Study is completed.

- To move through the NTP eHome Study, you will use the grey buttons or blue arrows at the bottom of each screen. You have the option of skipping questions and coming back to them later if you are unsure of the correct response using the buttons at the bottom of the screen.

- As appropriate, on each eHome Study screen you will see:
  - **First**: moves back to the first screen of the NTP eHome Study
  - **Previous**: takes you to the previous screen
  - **Next**: advances to the next screen
  - **Last**: moves to the last screen of the NTP eHome Study

- **Save and Exit**: saves your responses and exits the NTP eHome Study. You can come back to complete it later, and you will be on the screen with the first unanswered question.
- **Submit Test**: only seen on the last screen, the "Submit Test" button checks to be sure all questions are answered, and if so, submits the completed NTP eHome Study to RTI. If all questions are not answered, it will instruct you to answer the remaining questions.
- **Progress Bar**: shown at the bottom of the screen, the progress bar will fill with blue as you proceed through the eHome Study questions.

Do not click the "Back" or "Forward" buttons in your browser (green arrows in the top left corner of the screen).

Do not click the "X" (top right corner of the screen) to exit. If you click the "X" to exit, your responses will not be saved and you will have to re-enter them.

Always use the "Save and Exit" or "Submit Test" buttons to save and submit your eHome Study.

- If you must stop before you have completed the entire eHome Study, click on "Save and Exit" to save the answers you have entered so far. To re-enter the eHome Study later, follow the same steps as the first time you entered: go to the website, enter your FI ID, click "Login," confirm your name, and the program will automatically take you to the screen with the first unanswered question.
- You may change your answers at any time (even after you have clicked "Save and Exit"), up until you click "Submit Test."
- If you experience any difficulty accessing or completing the eHome Study, you should not click "Submit Test" until you have spoken with your FS. Once you submit the eHome Study, your answers are considered final and cannot be changed. However, you can "Save and Exit" as much as you like.
- To submit your eHome Study, click "Submit Test" on the final page and the program will check to see that you have answered all questions.
  - If you have not answered all of the questions, you will be taken back to the first unanswered question.
  - If you have answered all of the questions, you will see a confirmation screen asking if you are ready to submit your answers to RTI. Click "Yes," and your responses will be saved and submitted to RTI. Once you submit the eHome Study, you can no longer return to it.

III. ADDITIONAL PREPARATIONS FOR NSDUH TRAINING

In addition to completing the NTP eHome Study, there are specific project materials you must bring to training. To ensure you have all required items, use the following check list as you pack for training:

Items You Must Bring to Training:

____ 2012 NSDUH FI Manual and Computer Manual

____ 2012 Mental Health Surveillance Study FI Handbook

____ All required Headway Forms and documentation necessary to complete Section 2 of your I-9 Form. Forms are located in your Headway Employment Package, which you received in a separate shipment from Headway.

____ Two forms of identification required for the fingerprinting process: One must be a state or federally issued ID card (driver license or another Federal Government ID card). The other may be a Social Security card, military ID, voter registration card, passport or permanent resident card. You must bring the original documents, not copies.

IV. UPON ARRIVAL AT THE NSDUH TRAINING SITE

When checking into the hotel, ask the front desk for the location of **NSDUH Registration**. Go to the NSDUH Registration as soon as possible after you check in and drop off your bags in your hotel room.

Be sure to bring the following with you to NSDUH Registration:

____ All required Headway forms

____ Your travel itinerary with departure information

____ Appropriate ID for employment verification and fingerprinting (i.e., valid driver license and Social Security Card or passport)

While at NSDUH Registration, you will:

- Have your photo taken for your ID badge
- Complete necessary administrative forms
- Receive additional information about training
- Be fingerprinted for security purposes
- Turn in completed Headway forms

V. ADDITIONAL INFORMATION ABOUT NSDUH TRAINING

- The temperatures in training classrooms often vary so please dress in layers to help regulate your personal comfort.
- During training, FI Labs will be available to you in the evenings, which provide an opportunity to practice in any areas desired with trainers present to assist and answer questions. However, in the interest of strengthening your skills, your FS or trainers may **require** you to attend FI Lab.
- All FIs are required to undergo a certification at the end of training, where each FI works one-on-one with a trainer to complete a basic NSDUH screening and interview. Certifications occur after class on Days 5, 6 and 7.
- After training, every FI is required to complete a post-training teleconference with his/her FS and mentoring in the field by an FS or experienced FI. Your FS will schedule these important post-training activities. Additionally, FIs must complete several post-training courses independently using an interactive program called iLearning. You will learn more about this program at training.
- You will be compensated for the time spent on the training activities outlined in this memo (NTP eHome Study, FI Labs, certification, homework, post-training teleconference, and post-training iLearning courses). The training check you receive at the end of training will include payment for an additional 25.75 hours, intended to cover the non-classroom time spent on these activities.
- If you have any questions about the information contained in this NSDUH eHome Study package or any other project-related questions, please contact your FS.

We look forward to seeing you at NSDUH Training!

Appendix B: New-to-Project Electronic Home Study Exercises

2012 National Survey on Drug Use and Health
Home Study Exercises

This paper version of the NSDUH eHome Study is provided for your reference to use as needed while reviewing your manuals and completing the web-based exercises.

Please select a response for each question.

Section 1 – NSDUH FI Manual

Use your **NSDUH FI Manual** for reference to answer these questions. Select the best possible answer.

1. What agency sponsors the survey?

   a. National Center for Health Statistics
   b. National Institute on Drug Abuse
   c. Substance Abuse and Mental Health Services Administration
   d. Food and Drug Administration

2. Which of the following is **NOT** a goal of the NSDUH?

   a. To track trends in the use of alcohol, tobacco, and various types of drugs
   b. To provide accurate data on the level and patterns of alcohol, tobacco, and illegal substance use and abuse
   c. To identify those groups at high risk for substance use and abuse
   d. To assess the consequences of substance use and abuse
   e. To track an individual's patterns of substance use over time

3. NSDUH FIs should be available approximately 20 – 25 hours per week to conduct screening and interviewing during the data collection period.

   a. True
   b. False

4. Which of the following is your responsibility in the screening and interviewing process?

   a. Mailing a lead letter to each sample dwelling unit (SDU) that has a mailable address (your FS does this for your initial assignment)
   b. Locating (using the segment materials) and contacting SDUs
   c. Obtaining informed consent from a respondent (gaining permission from a parent/guardian before approaching a youth respondent)
   d. Transmitting the data to RTI on a daily basis
   e. All of the above
   f. a. and b. only
   g. b., c., and d. only

5. One very important requirement of your job is the proper treatment of the data, that is, keeping data completely confidential. Which information must you keep confidential?

   a. Answers provided during screening
   b. Answers provided during the interview
   c. Observed information from before the interview
   d. Observed information during or after the interview
   e. a. and c. only
   f. Any and all information you learn about the respondents

6. Group Quarters Units (GQUs) are generally any single living unit within a group quarters structure in which 10 or more unrelated persons reside.

   a. True
   b. False

7. What information does the Selected Dwelling Unit List provide?

   a. Telephone numbers for all selected respondents
   b. A list of housing units and group quarters units selected in the segment
   c. A list of all of the housing units and group quarters units found in the segment
   d. All 'next listed lines' that follow a Selected Dwelling Unit
   e. b. and d. only

8. Which of the following information is included on the iPAQ's **Select Case** screen?

   a. The RTI case identification number, referred to as the "Case ID number"
   b. The street address, or a physical description of the SDU and its general location
   c. The number of residents of the SDU
   d. All of the above
   e. a. and b. only

9. When do you make an entry in the Record of Calls (ROC)?

   a. Each time you discuss the SDU with your FS
   b. Each time you think about visiting the SDU
   c. Each time you attempt to contact the SDU
   d. Each time you actually speak with someone at the SDU
   e. a., c., and d. only
   f. c. and d. only

10. According to the NSDUH FI Manual, two productive time frames to visit SDUs are before 9am on weekend mornings and from Noon until 2pm during the week.

    a. True
    b. False

11. Who is an eligible screening respondent for the NSDUH?

    a. Any resident of the dwelling unit (DU)
    b. Any adult (age 18 or over) who answers the door
    c. An adult (age 18 or over) resident of the DU
    d. Anyone that lives on the street

12. You must always wear your RTI photo ID badge when working on the NSDUH in the field.

    a. True
    b. False

13. According to the NSDUH FI Manual, two steps you can take to reduce refusals to participation include being able to explain the purpose of the study and believing in yourself.

    a. True
    b. False

14. The screening process includes questions about:

    a. The number of people age 12 and over who will live at the SDU for most of the quarter
    b. The correct address
    c. The number of residents in the household who take licit and illicit drugs
    d. Age, relationship, gender, Hispanic origin, race, and military status
    e. b. and c. only
    f. a., b., and d. only

15. It is possible for the screening process to identify:

    a. One eligible household member
    b. Two eligible household members
    c. No one eligible in the household
    d. Either a., b., or c.

16. After entering physical features data, which screen should be displayed on your iPAQ when you knock on the door of the SDU?

    a. Select Case Screen
    b. Study Introduction Screen
    c. Record of Calls Screen
    d. None of the above

17. You must read the Informed Consent screen on the iPAQ and give a Study Description to every Screening Respondent.

    a. True
    b. False

18. You should always attempt to complete the NSDUH interview:

    a. Immediately after screening
    b. At a later date, to give the respondent time to prepare
    c. With other household members in the same room, so the respondent feels more at home
    d. With a parent or guardian in the same room for minor respondents
    e. In complete privacy
    f. a. and d. only
    g. b. and c. only
    h. a. and e. only

19. A good response to a parent who hesitates to let his child participate in the study because he thinks his child has not used drugs is:

    a. I'll mail you a copy of your child's answers so you can discuss them together.
    b. If your child turns out not to use drugs, we'll throw the data out.
    c. Your child looks like he has had plenty of experience using drugs. I'm sure he'll be a great respondent.
    d. There are other topics included besides drugs. Knowing the opinions and experiences of your child is important as well.

20. If a respondent doesn't understand a question, you should rephrase it in your own words until the respondent comes up with an answer.

    a. True
    b. False

21. Which of the following is **not** an acceptable probe?

    a. To repeat the question
    b. To pause
    c. To repeat the answer choices
    d. To suggest answers
    e. To use neutral questions or statements

22. Respondents will be offered a cash incentive of $30 for completing the entire NSDUH interview.

    a. True
    b. False

23. What is the **minimum** number of times you are **required** to report to your FS by phone?

    a. At least twice per week
    b. At least twice per month
    c. At least once per week
    d. At least once per month

24. The NSDUH's deadline for transmitting your weekly ePTE summary data from the iPAQ is 11:30pm every Saturday night (Eastern Standard Time).

    a. True
    b. False

25. For certain non-interview screening codes, you are required to obtain verification information about the contact person. What information must you record in the iPAQ?

    a. First name, last name, and phone number
    b. First name and phone number
    c. Phone number only
    d. None of the above

26. What time period does the ePTE cover?

    a. 2-week period
    b. 1-day period
    c. 1-week period
    d. 1-month period

27. Before leaving your home to go work in the field, if the time and date displayed on the iPAQ are **not** correct, you should:

    a. Wait and work another day
    b. Call your FS
    c. Connect the iPAQ to the laptop and transmit
    d. Disregard the time and date and go to work

28. NSDUH FIs are allowed to gather screening information from a neighbor after three failed attempts at contacting the residents of the SDU.

    a. True
    b. False

29. If you are conducting two interviews at the same DU, you can use the same Reference Date Calendar for both respondents.

    a. True
    b. False

30. NSDUH protocol requires that you always plug in and offer the headphones to each interview respondent.

    a. True
    b. False

31. What SDU Characteristic data should you enter for an SDU within a senior housing apartment building with 82 units?

    a. Multi-unit, 50+ units
    b. Senior Housing/Assisted Living
    c. Other GQU
    d. a. and b. only

32. NSDUH's missed dwelling unit procedures require FIs to check for missed DUs at every dwelling unit listed in the segment.

    a. True
    b. False

33. When must completed Quality Control forms be mailed to RTI?

    a. On a weekly basis
    b. After accumulating 10 or more completed forms
    c. Within 24 hours of the completion of the interview
    d. Never – the forms are for your records only
    e. None of the above

34. You should **not** sign the Interview Payment Receipt during the ACASI portion of an interview; you should always wait until you are prompted to do so by the laptop.

    a. True
    b. False

Section 2 – NSDUH FI Computer Manual

Use your **NSDUH FI Computer Manual** for reference to answer these questions. Select the best possible answer.

35. Which of the following is an advantage to using computer-assisted personal interviewing (CAPI)?

    a. Identifies inconsistencies in responses to critical items and lets you resolve them in the best way: with direct and immediate input from the respondent
    b. Allows for intricate question and skip patterns based on entered data
    c. Saves time and project resources by combining both interviewing and data entry
    d. All of the above

36. To "tap" on the iPAQ, you can use the special iPAQ stylus (pen) or any regular pen.

    a. True
    b. False

37. Transmission of CAI interview data and iPAQ screening and ROC data is conducted via a single transmission from the laptop.

    a. True
    b. False

38. In the iPAQ screening program, text displayed in red, capital letters is text to be read to the respondent.

    a. True
    b. False

39. From the CAI Manager, you can:

    a. Send e-mail
    b. Start a NSDUH interview
    c. Transmit completed interview data to RTI
    d. Read e-mail from RTI
    e. Submit ePTE reports
    f. b., c., d., and e. only

40. The 3-letter code you need to move from the ACASI section back into the CAPI interview is:

    a. CAI
    b. RTI
    c. Your initials
    d. To be distributed at training

41. You are allowed to use the Touchpad on the laptop during an actual CAI interview.

    a. True
    b. False

42. To clean the laptop screen, you should:

    a. Use a cloth dampened with water only
    b. Use a cloth dampened with soap and water
    c. Spray the screen with a cleaning solution
    d. None of the above

43. If the CAI Manager is "frozen" and won't accept any data during the interview:

    a. You may have accidentally entered an extra space in the answer field
    b. The CAI program is too cold
    c. The title bar at the top of the screen is light blue/gray and you need to press [Alt] [Tab]
    d. a. and c. only

44. If you are in a respondent's home and cannot complete the screening or interview because of a technical problem, you should:

    a. Call your FS immediately
    b. Call Technical Support immediately
    c. Break off the screening or interview and come back when your equipment works
    d. None of the above

Internet Information Questions

Please answer the following questions concerning your internet availability and access. These answers will not be a part of your home study score and will only be used for information purposes.

45. In order to complete the electronic home study, where did you access the internet?

    a. Home
    b. School
    c. A workplace
    d. A friend, neighbor, or relative's house
    e. A public library, community center, internet café, coffee shop, or some other place with free internet access
    f. A store, internet café, or some other place where you pay for access to the internet
    g. A Smartphone, such as a Blackberry or iPhone

46. What was your internet speed? (If unsure, take your best guess.)

    a. High speed (e.g. cable, DSL, broadband, etc)
    b. Telephone modem (e.g. Dial-up)

47. What type of computer did you use?

    a. PC (most likely running Windows, Windows XP or Windows Vista)
    b. Mac (MacBook laptop, iMac, etc.)

48. Did you have any difficulties accessing or completing the electronic home study?

    a. Yes
    b. No

Appendix C: Veteran FI Training iLearning Courses Memo

RTI INTERNATIONAL | Memorandum

DATE: **November 28, 2011**
TO: **2012 NSDUH Veteran Field Interviewers**
FROM: **Ilona Johnson, National Field Director**
RE: **2012 NSDUH Veteran FI Training iLearning Courses**

The 2012 NSDUH Veteran FI Training Program consists of several iLearning courses to be completed by the deadlines provided in this memo, a review of the 2012 manuals, and a one-day FS Team Meeting hosted in early January 2012. Enclosed are the materials needed to complete these important training tasks in preparation for 2012. Please read this memo carefully and review the contents of this shipment before beginning your training assignment. In addition to this memo, your shipment includes:

- 2012 NSDUH Veteran Training iLearning Courses CD
- 2012 Mental Health Surveillance Study (MHSS) FI Handbook
- 2012 FI Manual and FI Computer Manual

Begin the training tasks outlined in the chart on Pages 1 and 2 as soon as possible to ensure there are no technical problems. Use the columns provided to track your progress.

If you have any questions or are missing any items from this shipment, please contact your FS immediately. If you have any technical questions or problems, please contact your FS first, and if needed, your FS will instruct you to contact NSDUH Technical Support at 877-419-1768.

Complete the following 2012 Veteran Training tasks:

2012 Veteran Training iLearning Courses

After viewing the 2012 Veteran Training Introduction Video, complete and transmit the iLearning Courses in the order listed below **by 11:59 PM Eastern Standard Time according to the deadline**.

Use the **approximate** course length provided to help plan your schedule. You are encouraged to complete all courses prior to the deadlines.

Deadline	Course Name	Date Completed	Date Transmitted
No later than Wednesday, December 7, 2011	"2012 NSDUH Updates" (~ 45 minutes) "2012 CIPSEA Training" (~ 45 minutes)		
No later than Wednesday, December 14, 2011	"2012 Mental Health Surveillance Study" (~ 30 minutes) "2012 Habits For Success" (~ 45 minutes)		

As a refresher on key project procedures, you will complete the "2012 Mental Health Surveillance Study" and "2012 Habits For Success" iLearning courses prior to Quarters 2, 3 and 4 in 2012. For this reason, store your iLearning CD in a safe location at home so it can be easily accessed when needed.

***NOTE:** Do not complete the "FS Team Meeting Feedback" course at this time. This course will be completed following your FS Team Meeting in January. You will receive more details on this course during your team meeting.

Complete the following 2012 Veteran Training tasks, continued:		Task Complete?
2012 Mental Health Surveillance Study (MHSS) FI Handbook	The Mental Health Surveillance Study will continue as part of the 2012 NSDUH. Carefully review the updated 2012 MHSS FI Handbook prior to your FS Team Meeting in January and refer to it throughout the year when questions arise.	
2012 FI Manual and FI Computer	The FI Manual and FI Computer Manual have been updated for 2012 and you should carefully review the changes detailed in the chart beginning on Page 6 of this memo. In addition, you must read the manual pages listed below due to the number and importance of the updates made to these sections. This task must be completed prior to attending your FS Team Meeting in January. ***FI Manual Chapter 6 – Conducting Screening:*** PAGES 6-27 to 6-29 (Section 6.5) and PAGES 6-39 to 6-41 (Section 6.8) ***FI Manual Chapter 8 – The NSDUH Interview:*** PAGES 8-13 to 8-15 (Sections 8.6 and 8.7) and PAGES 8-23 to 8-26 (Section 8.10) ***FI Manual Chapter 9 – Documenting and Reporting Procedures:*** ALL PAGES ***FI Manual Appendix C – Refusal/Controlled Access Letters:*** PAGES C-18 to C-30 ***FI Computer Manual Appendix B — Transmission Connection Wizard:*** ALL PAGES [NOTE: This appendix has been added for 2012 to detail the use of the new Transmission Connection Wizard software. You will receive more information about when this software will be released and available for use at your team meeting.]	

Key items to remember while completing the 2012 Veteran Training iLearning Courses:

- Click the "iLearning" button once to begin an iLearning course. As needed, refer to **Appendix F** in your **FI Manual** for information on beginning and exiting an iLearning course.
- **Connect your laptop to electrical power** (versus running on battery power) and **use the headphones** to ensure the best audio quality.
- With your laptop at the CAI Manager, the **first** time you load the iLearning CD a message will display indicating your computer will be optimized for iLearning. Do not touch the computer once you see this message. The optimization process should only take 1-2 minutes. Once the optimization process is finished, another message will display to confirm the process has been completed. Click "OK" to continue and you will return to the CAI Manager.
- **Listen to the audio in its entirety** before viewing a video or proceeding to the next screen. If you move too quickly through the lesson portion, you may miss important points which might be covered in the assessment questions. Additionally, moving too quickly may cause the audio to load improperly or skip.
- Each course includes an assessment portion with 5-10 questions. Each course is scored separately. You must score at least 80% to pass the course. To ensure understanding, your FS will review any missed questions with you within one week following the course deadline. Any FI who does not achieve a passing score will be placed on probation and required to complete additional training before beginning their Quarter 1 assignment. Keep in mind you will have two opportunities to answer each question, and **the expectation is that EVERY FI will pass each course**. (Note: The FS Team Meeting Feedback course does not include an assessment.)
- If you have any questions regarding the status of your iLearning courses or whether the results have been received at RTI, **refer to the text located above the course list** in the menu on the left hand side of the iLearning screen. After selecting a course by clicking on the title once, you will see the status. For any additional questions, refer to **Appendix F** or contact your FS.
- **In order to continue your role as a NSDUH FI in 2012, you must successfully complete and transmit the iLearning courses according to the deadlines outlined in this memo.** You are encouraged to complete all iLearning courses well ahead of the deadlines. There is no need to complete the courses at different times and if possible, you could complete multiple courses in one sitting. If an individual iLearning course deadline is missed a verbal warning will be issued, so it is important to manage your time effectively to meet these deadlines.

Charge time spent completing and transmitting the 2012 Veteran Training iLearning courses, and reviewing the 2012 FI Manual, FI Computer Manual, and Mental Health Study FI Handbook to **0212800–001.105.002 under the 'Training' column of your ePTE with appropriate notes**. The total time for completing these tasks is expected to be less than 6 hours.

For your reference, an overview of the 2012 NSDUH Veteran FI Training schedule and the 2012 NSDUH Materials "Keep" List are provided on Page 4. When finished with ALL Quarter 4 2011 work, keep the materials included on this list for use in 2012 and recycle or throw away any materials NOT listed. You should receive your 2012 bulk supplies on December 15, 2011, so it is important to discard any unusable materials before unpacking your bulk supplies to avoid any confusion.

Thank you for your attention to these details.
We hope you enjoy this year's Veteran Training program!

NSDUH Materials "Keep" List

Please keep any of the below materials you have for use during 2012 NSDUH data collection.

- Advance Balance Sheets
- Certificates of Participation
- Headway Expense Reports
- Headway Expense Report Working Copies
- Newspaper Articles
- NSDUH Short Reports
- Other Language Introduction Cards
- Paper PT&Es
- Plain Envelopes
- Receipts for Escort Services
- Segment Access Documentation Forms
- Spanish Cards

Any materials that are not listed above should be recycled [or thrown away] and cannot be used on the 2012 NSDUH.

2012 NSDUH Veteran FI Training Schedule	Date
Receive 2012 NSDUH Veteran FI Training iLearning CD, FI Manual and MHSS FI Handbook	Tuesday, November 29, 2011
Complete & transmit these iLearning Courses: "2012 NSDUH Updates" "2012 CIPSEA Training"	No later than Wednesday, December 7, 2011
Complete & transmit these iLearning Courses: "2012 Mental Health Surveillance Study" "2012 Habits For Success"	No later than Wednesday, December 14, 2011
Receive 2012 bulk supplies shipment, including the 2012 CAI & iPAQ Program Updates CD and instructions	Thursday, December 15, 2011
Complete installation of the 2012 CAI & iPAQ Programs (once you are finished with all Quarter 4 2011 work)	December 15 - 28, 2011
FS Team Meeting (Session A and B)	January 4 or 6, 2012 (9:00 am – 4:00 pm)

2012 FI Manual Changes		
Item	**Location**	**Change**
***Section 1.2* – Project Organization**	Pg. 1-1	Updated reference to Headway Workforce Solutions (to reflect the recent name change).
***Section 1.3.1* – Brief History of NSDUH**	Pg. 1-3	Updated the NSDUH goals to match the objectives listed on the Q&A brochure and NSDUH Respondent website.
Exhibits 1.3 through 1.6	Pgs. 1-10 thru 1-15	Re-ordered Exhibits 1.3, 1.4, 1.5, and 1.6 to reflect the order in which each is mentioned in the manual text.
***Exhibit 1.5* – Project Staff Information**	Pg. 1-14	Updated project staff contact information.
***Exhibit 1.6* – FI Manual Navigation Functions**	Pg. 1-15	Added a second method for searching the FI Manual using [**Ctrl**] + [**f**].
***Review of Chapter 1*– Questions to Ask Yourself**	Pg. 1-18	Added space after each question to record answers and other notes. This change was made at the end of every chapter in the manual.
***Exhibit 2.2* – Job Description for a NSDUH Field Interviewer**	Pg. 2-5	Updated the project number and made minor wording edits in the requirements for screening/interviewing activity.
***Section 2.5*– Importance of Confidentiality**	Pg. 2-7	Provided examples of identifying information (such as DU addresses, respondent names or phone numbers) that should not be written on any materials or in emails to your FS.
***Section 2.8* – Materials, Supplies, and Equipment**	Pg. 2-9	Added instructions to inventory materials on a weekly basis and before your scheduled FS conference call. Included a reminder to store materials carefully to prevent damage since some materials, such as the Showcard Booklet, are expected to last the entire year.
***Review of Chapter 2*– Summary**	Pg. 2-11	Added a reminder that properly caring for your equipment and materials is an important field interviewer task and part of being cost-effective.
***Exhibit 32* – County Locator Map**	Pg. 3-4	Updated the project number shown on the map. The remaining maps in Chapter 3 were updated similarly.
***Section 3.3.1* – Housing Units**	Pg. 3-23	Added descriptions for DU #8 and DU#20 in the examples at the bottom of the page.
***Section 3.6.1* – Adding Missed Housing Units**	Pg. 3-38	Updated instructions on what to record if a specific address is not available for a missed housing unit. After entering a brief description of the unit in the Street Number field, you should record the street number and street name of the original SDU where the unit was found in the Street Name field. A reminder to use a simplified description in the Street Number field and an example were also added.
***Section 3.6.2* – Reconciling Missed DUs**	Pg. 3-40	Added a reminder to read the iPAQ screens carefully when reconciling missed DUs.
***Section 4.2* – Assignment Materials**	Pg. 4-2	Updated text on where to include notes about specific cases. To ensure confidentiality, any notes should be documented in the iPAQ.

2012 FI Manual Changes, continued		
Item	**Location**	**Change**
***Section 4.5* – Lead Letters**	Pg. 4-12	Revised instructions for completing the lead letter. You are required to print the date and your name neatly on the letters to personalize them.
***Section 4.7* – Initial Approach**	Pg. 4-14	Added a reminder to provide a copy of the lead letter if the respondent doesn't remember receiving one.
***Section 4.8* – Your Introduction**	Pg. 4-14	Added further explanation that while the Study Introduction screen does not have to be read verbatim, it is worded in a concise and direct manner. Using this text as much as possible in your introduction helps to ensure all four required points are included.
***Section 4.9* – Eligible Screening Respondent and Address Verification**	Pg. 4-15	Revised text to reflect that telephone screenings are not allowed under any circumstances on NSDUH.
***Section 4.9* – Eligible Screening Respondent and Address Verification**	Pg. 4-15	Added a reminder to not leave a "Sorry I Missed You" card where it is easily seen because this may indicate to others the residents are away, leading to possible theft.
***Section 4.12* – Handling Controlled Access Situations**	Pgs. 4-25 and 4-26	Added a brief description of the Door Person Letter and its proper usage.
***Section 4.12* – Handling Controlled Access Situations**	Pg. 4-26	Added text indicating copies of the Controlled Access Letters are included in Appendix C.
***Exhibit 4.6* – Door Person Letter**	Pg. 4-27	Added the Door Person Letter as a new exhibit. From here forward, the exhibit numbering has been updated to reflect the new exhibit.
***Exhibit 4.9* – Controlled Access Issues and Possible Solutions**	Pg. 4-32	Added a note in the College/Student Housing section to talk with your FS first when setting up screenings/interviews in a common area.
***Exhibit 5.6* – Answering Questions**	Pg. 5-11	Revised and simplified possible responses to several questions.
***Section 5.6* – Overcoming Objections**	Pg. 5-20	Clarified that a refusal letter is not sent when the refusal is situational (i.e. ill, house too messy) or when "Need to discuss with FS" is selected.
***Section 5.10.1* – The Home**	Pg. 5-23	Added reminders about inventorying materials and to only request materials as needed, as they are very costly to produce.
***Exhibit 5.14* – Working Safely in Winter Weather**	Pg. 5-34	Added a new exhibit with tips for working in winter weather.
***Section 6.4.6* – Completing the Housing Unit Roster**	Pg. 6-13	Added a reminder that the text in parenthesis on the Total SDU Members screen should not be read at ALL households, but only if you feel the SR needs additional help understanding who to roster.
***Section 6.4.6* – Completing the Housing Unit Roster**	Pg. 6-16	Added text indicating that age ranges should only be entered when absolutely necessary, as an accurate age is needed for analysis of the data.

2012 FI Manual Changes, continued		
Item	**Location**	**Change**
***Section 6.5.1* – Entering Verification Data**	Pg. 6-27	Added a reminder about confirming the phone number with the respondent to ensure you have entered it correctly.
***Section 6.5.1* – Entering Verification Data**	Pg. 6-28	Added instruction to never prompt the respondent for their phone number by leading with the area code when obtaining verification information. The respondent should provide the entire phone number, including the area code.
***Section 6.5.1* – Entering Verification Data**	Pg. 6-29	Added a note that if a name is not available for the contact person, to include a note to the verification caller about the situation. Also, if a translator was used to conduct the screening, include a note documenting the use of a translator and the language spoken, if known.
***Exhibit 6.2* – Editing Address Protocol Chart**	Pg. 6-40	Updated instructions on the first page of the Editing Address Protocol Chart.
***Section 6.10* – Physical Features**	Pg. 6-45	Updated instructions on coding multi-units. If you encounter no controlled access barrier during your initial visit to a multi-unit building or complex, code "None" as the controlled access type for the units you actually visited.
***Section 6.20* – Call Distribution**	Pg. 6-54	Updated the name of the Call Distribution feature to match the iPAQ. This update was made throughout the manual.
***Section 7.3.2* – Interview Result Codes**	Pg. 7-5	Clarified that two ROCs should be entered for breakoffs where an appointment was scheduled to complete the interview at a later time. One ROC should be entered to document the breakoff, and another for the interview appointment.
***Section 7.5* – Obtaining Interview Participation**	Pg. 7-11	Added a reminder to be present and available during the ACASI to provide the pillcards to the respondent and in case any problems arise.
***Section 7.7.1* – Choosing a Location**	Pg. 7-26	Added a reminder to never place the laptop on your lap or the respondent's lap. Doing so blocks the air vents and the laptop may become hot enough to harm skin and damage the equipment.
***Section 7.7.1* – Choosing a Location**	Pg. 7-27	Added text about not conducting interviews in locations where there is a possible risk of damage to the equipment.
***Section 7.7.3* – Setting Up the Computer**	Pg. 7-29	Added a reminder about only using the NSDUH provided headphones for the interview.
***Section 8.6* – Use of Showcards and Pillcards**	Pg. 8-13	Added text about reading the CAI screen to the respondent to indicate the computer will tell them to ask you for some cards.
***Section 8.7* – Completing the Reference Date Calendar**	Pg. 8-14	Indicated that the month should be written in words on the calendar and 4 digits should be used for the year. Also added a reminder not to leave or offer the calendar from a breakoff interview. For breakoffs, no calendar will be used once the interview is resumed.
***Section 8.8* – FI Responsibilities During the ACASI Portion**	Pg. 8-17	Replaced images during this section with actual CAI screen shots.
***Section 8.8* – FI Responsibilities During the ACASI Portion**	Pg. 8-21	Updated suggestions on how to use the time during the ACASI and added a reminder to remain attentive while completing the Data Quality Knowledge Notebook.

2012 FI Manual Changes, continued		
Item	**Location**	**Change**
***Section 8.10* – Respondent Difficulties**	Pg. 8-24	Included additional instructions on how to handle interviewing blind, deaf, and other physically impaired respondents.
***Section 9.1* – Introduction**	Pg. 9-1	Added a reminder about mailing completed QC Forms to RTI within 24 hours of interview completion.
***Section 9.3* – Entering Weekly PT&E Summary Data in iPAQ**	Pg. 9-4	Updated the task number here and throughout the manual.
***Section 9.4* – Conference Call with Field Supervisor**	Pg. 9-7	Added reminders on preparing for conference calls with your FS.
***Exhibit 9.2* – Conference Call Checklist**	Pg. 9-9	Added a new exhibit to provide a checklist of advance preparations, etc. for conference calls with your FS.
***Section 11.5.2* – Default Value Settings**	Pg. 11-6	The project number for 2012 data collection is 0212800-001.106.002. This update has been made throughout the manual.
***Section 11.5.4* – Detail Form**	Pg. 11-15	Added a reminder that mileage for separate trips to the bank to obtain incentive cash is not an allowable charge. Be efficient in planning trips so you can stop by the bank on the way to your segment, etc.
***Section 11.5.4* – Detail Form**	Pg. 11-16	Added a reminder that parking tickets, traffic tickets and other fines are not reimbursable expenses. In general, ATM fees are also not reimbursable. Added that a summary of incentive monies paid, as well as an explanation for any case in which the respondent refused to accept the incentive (including the CaseID) should be included in the Notes section of the ePTE.
***Section 11.7.3* – Repaying the Incentive Advance Balance**	Pg. 11-29	Updated text about repaying the incentive advance to reflect current procedures. Any FI on inactive status for 30 days or more must repay the advance in full immediately upon the change to inactive status. Check with your FS for detailed instructions on submitting payments to Headway.
***Exhibit 11.5* – Incentive Advance Agreement**	Pg. 11-30	Updated the Headway Incentive and Travel Advance Agreement exhibit.
***Exhibit 11.10* – Travel Assignment Information**	Pg. 11-40	Updated this exhibit to reflect current procedures.
***Section 12.2* – Materials and Equipment**	Pg. 12-2	Added the Door Person Letter to the list of additional materials for field reference/use.
***Section 12.5* – Weekly Tasks**	Pg. 12-6	Added the task of inventorying your materials weekly.
***Section 12.7* – End of Quarter Tasks**	Pg. 12-7	Updated the end of quarter iPAQ transmission instructions to reflect current procedures. First you should transmit twice from your iPAQ back to back. After waiting 2 hours, transmit again to pick up your next quarter's work.

2012 FI Manual Changes, continued		
Item	**Location**	**Change**
***Appendix A* – Completing Paper PT&Es**	Pg. A-4 & A-5	Added text indicating that mileage for trips to the bank, parking tickets, traffic tickets, or other fines are not reimbursable expenses. Added a reminder that a summary of incentives paid should be included in the notes section of the ePTE.
***Appendix C* – Refusal/Controlled Access Letters**	N/A	Added the Controlled Access Letters to this Appendix for easy reference and updated the Appendix title to reflect this addition.
***Appendix D* – NSDUH Respondent Website Content**	N/A	Updated the screen shots with the latest version of the NSDUH Respondent website.
***FI Computer Manual*: Section 3.3.5 – Reset**	Pg. C3-11	Added text indicating that you should reset your iPAQ before plugging it in to charge at the end of the day and before going out to work the next day to improve performance.
***FI Computer Manual*: Section 4.2.1 – The Left Side**	Pg. C4-2	Added a reminder to never place the laptop on your lap or the respondent's lap. Always place the laptop on a flat, sturdy surface. This update was made throughout the manual.
***FI Computer Manual:* Appendix B – Transmission Connection Wizard**	N/A	Added this Appendix to detail the use of the new Transmission Connection Wizard software that allows for transmissions via Ethernet or wireless internet connections in addition to dial-up. Carefully read this Appendix prior to attending your FS Team Meeting. You will receive more information about when this software will be released and available for use at your team meeting. Until that time, continue to use the standard dial-up transmission process (via an analog telephone line) described in Chapter 6 of the FI Computer Manual.

Appendix D: Data Collection Preparations Memo

RTI INTERNATIONAL | Memorandum

DATE: **December 14, 2011**

TO: **NSDUH Veteran Field Interviewers**

FROM: **Ilona Johnson, National Field Director**

RE: **2012 NSDUH Data Collection Preparations**

Enclosed are the materials designed to assist you in loading the 2012 CAI and iPAQ programs on your project equipment and completing other important tasks prior to the start of Quarter 1 data collection in January. Please read this memo and review the contents of this shipment carefully before beginning your preparations. In addition to this memo, your shipment includes:

- **2012 NSDUH CAI & iPAQ Update Instructions**
- **2012 NSDUH CAI & iPAQ Updates CD**
- **2012 NSDUH Veteran FI Bulk Supplies**

Begin the preparations outlined below as soon as possible after receiving this shipment to ensure there are no technical problems with your CD or equipment. If you are missing any items in this shipment, please contact your field supervisor immediately.

Complete the following in preparation for 2012 Data Collection:	Deadline
Install the 2012 CAI and iPAQ Programs: • Only after you have completed all of your Quarter 4 2011 work, install the 2012 programs following the "2012 NSDUH CAI & iPAQ Update Instructions." After installation, you will transmit a practice break-off screening and interview to confirm the updates were installed correctly.	No later than December 28, 2011
Recycle or discard any 2011 materials not listed on the NSDUH Materials "Keep" List. A copy of the "Keep" List can be found with your bulk supplies and the iLearning courses memo sent in November: • To avoid confusion, it is important to discard any unusable 2011 materials prior to unpacking the 2012 bulk supplies included with this shipment. • FYI—Your bulk supply shipment also includes an Ethernet cable and an extra headset to use as back-up. Keep the extra headset in your laptop bag or car so it is easily available if needed. Store the Ethernet cable in a safe location at home until you receive further instructions.	No later than January 2, 2012

In order to continue in your role as a NSDUH FI in 2012 and begin your Quarter 1 assignment, you must:

1. Pass all 2012 Veteran Training iLearning courses
2. Successfully install the 2012 CAI and iPAQ programs; and
3. Attend your FS Team Meeting in early January.

Quarter 1 cases will be released for downloading to your iPAQ on January 4, 2012 (for FIs attending the Session A FS Team Meeting) or January 6, 2012 (for FIs attending the Session B FS Team Meeting). If the first two requirements above have not been met, you will not be able to pick up your Quarter 1 cases.

<u>You must not begin work on your Quarter 1 cases until after attending your assigned FS Team Meeting in January.</u>

After installing the 2012 CAI and iPAQ programs and a successful transmission of a practice break-off screening and interview, you will need to update the project number for 2012 data collection in your ePTE Default Values. To update the project number, refer to the instructions below:

1. From the CAI Manager, click "PTE Entry." At the Entry Form, press [Alt] [t] (for Tools) and then [v] (for Update Default Values).
2. Next, the System Default Value Setting Form appears. In the Project Number field, enter **0212800** in the first field and **001.106.002** in the field after the "dash". Check your entry carefully before continuing.
3. After you make this change, click "Save." Once you have saved the form, you can exit by clicking "Exit." At the Entry Form, click "Exit" to return to the CAI Manger.

Charge your time for installing the 2012 CAI and iPAQ programs to **0212800–001.105.001** under the 'Other' column of your ePTE with appropriate notes. Refer to the "2012 NSDUH CAI & iPAQ Update Instructions" for additional information on charging your time.

Thank you for your attention to these details and for your continued commitment to NSDUH. We look forward to seeing you in January. If you have any questions or are missing any items in this shipment, please contact your field supervisor.

Thank you and good luck in 2012!

Appendix E: Verification Scripts

Verification scripts have been removed.

Appendix F: U.S. Census Bureau Industry and Occupation Coding Report

Industry and Occupation Coding

Overview

Toward the end of the National Survey on Drug Use and Health (NSDUH) questionnaire, the interviewer asked each respondent a series of questions to obtain details about the respondent's employment, including the type of business or industry and the main duties performed in the job. In 2012, the work of assigning industry and occupation codes for each respondent was completed by the National Processing Center (NPC) of the U.S. Census Bureau through an interagency agreement between the Substance Abuse and Mental Health Services Administration (SAMHSA) and the U.S. Census Bureau.

Process

RTI sent compiled industry and occupation questionnaire data to the NPC in four separate deliveries, one each quarter. NPC coders determined both an industry and an occupation code for each record; each code was determined at the four-digit level of detail. Coders used the U.S. Census Bureau 2007 standard industry and occupation classification coding system to assign the codes, meaning they used the 2002 North American Industry Classification System (NAICS) for industry coding and the 2000 Standard Occupational Classification (SOC) system for occupation coding.

Two different coders assigned the codes for each record. During the second coding, if the first and second codes did not agree, the second coder reconciled the discrepancy and assigned the final code. In some instances, cases were referred to a third party for assignment of a final code. The NPC then returned the codes to RTI for inclusion in the final NSDUH results. The NPC ensured that quality control measures were in place and adhered to, and it provided feedback regularly on production and error rates to coding staff.

Results

The NPC sent SAMHSA progress reports that included production rates per hour and numbers and percentages of codes requiring reconciliation separately for industry and occupation codes. Based on those reports, Tables F.1 through F.3 display the production information for the NPC coding process. Table F.1 provides the total number of completed interviews by quarter as well as the number of interviews containing industry and occupation data by quarter. Table F.2 contains the coding production results by quarter, while Table F.3 shows the production rates for each quarter.

Table F.1 2012 NSDUH Industry and Occupation Coding Production Report for the National Processing Center, U.S. Census Bureau—Completed Interviews, by Quarter

	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Completed Interviews	16,361	18,180	17,867	16,154	68,562*
Interviews with Industry and Occupation Data	9,629	10,732	10,877	9,704	40,942

*Completed interviews that were delivered to the U.S. Census Bureau throughout the year have not gone through the data cleaning and editing process; thus, the total is higher than the final number of completed interviews for the year.

Table F.2 2012 NSDUH Industry and Occupation Coding Production Report for the National Processing Center, U.S. Census Bureau—Production Results, by Quarter

	Quarter 1		Quarter 2		Quarter 3		Quarter 4		Total	
	Count	%	Count	%	Count	%	Count	%	Count	%
Total Coded	9,629	100.0	10,732	100.0	10,877	100.0	9,704	100.0	40,942	100.0
Total Verified	9,629	100.0	10,732	100.0	10,877	100.0	9,704	100.0	40,942	100.0
Industry Codes Requiring Reconciliation	556	5.8	568	5.3	759	7.0	772	8.0	2,655	6.5
Occupation Codes Requiring Reconciliation	1,054	11.0	964	9.0	1,201	11.0	1,223	12.6	4,442	10.8
Total Referred Cases	1,227	12.7	1,152	10.7	1,442	13.3	1,427	14.7	5,248	12.8

Total Coded: Codes assigned by first coder.

Total Verified: Codes assigned and confirmed by second coder.

Codes Requiring Reconciliation: First and second codes did not match. Second coder reconciled and assigned final code.

Total Referred Cases: Second coder could not reconcile case. Final code assigned by third-party coder using additional resources to resolve discrepancy.

Table F.3 2012 NSDUH Industry and Occupation Coding Production Report for the National Processing Center, U.S. Census Bureau—Production Rates, by Quarter

	Number per Hour				Average Number per Hour
	Quarter 1	**Quarter 2**	**Quarter 3**	**Quarter 4**	**Total**
Coding Production Rates	93.9	87.3	104.6	104.1	97.5
Coding Verification Rates	88.3	100.3	108.2	110.9	101.9
Problem Referral Rates	20.5	21.5	28.3	24.5	23.7

Coding Production: Codes assigned by first coder.

Coding Verification: Codes assigned and confirmed by second coder.

Problem Referral: Second coder could not reconcile case. Final code assigned by third-party coder using additional resources to resolve discrepancy.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

EDITING AND IMPUTATION REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.107.006.007
Deliverable No. 41

Authors:

Jeff Laufenberg
Larry Kroutil
Peter Frechtel
Lisa Carpenter
Susan Edwards
Glynis Ewing
Kristen Gulledge
Wafa Handley
Peilan Martin
Andrew Moore
Victoria Scott

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

April 2014

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

EDITING AND IMPUTATION REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.107.006.007
Deliverable No. 41

Authors:

Jeff Laufenberg
Larry Kroutil
Peter Frechtel
Lisa Carpenter
Susan Edwards
Glynis Ewing
Kristen Gulledge
Wafa Handley
Peilan Martin
Andrew Moore
Victoria Scott

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

April 2014

Acknowledgments

This report would not be possible without the guidance and input of staff from the Substance Abuse and Mental Health Services Administration (SAMHSA). In particular, Jonaki Bose and Grace O'Neill provided helpful feedback. Special thanks are also due to several current and former RTI International (a trade name of Research Triangle Institute) staff members. Ralph Folsom, Eric Grau (formerly with RTI), and Avinash Singh (formerly with RTI) co-developed the predictive mean neighborhood (PMN) methodology. Claudia Clark edited the report, Valerie Garner formatted the report, and Teresa Bass provided web production.

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Chapter **Page**

1. Introduction 1
   - 1.1 Organization of the NSDUH Questionnaire and Overview of Content 1
   - 1.2 Overall Data Quality Issues with CAI 6
   - 1.3 Organization of this Report 7
   - 1.4 Changes from the 2011 Survey to the 2012 Survey 8
     - 1.4.1 Changes to Accommodate Atypical Cases 9
     - 1.4.2 Changes to Streamline Imputation Procedures 9
     - 1.4.3 Changes Involving Minor Corrections and Recodes 10

2. Procedures and General Principles for Editing the Computer-Assisted Data 13
   - 2.1 Introduction 13
   - 2.2 Initial Processing of Transmitted NSDUH Interview Data 13
     - 2.2.1 Assignment of Standard NSDUH Missing Data Codes 14
     - 2.2.2 Remapping of Reponses to "Enter All that Apply" Questions 15
     - 2.2.3 Identification of Usable Cases 16
     - 2.2.4 Creation of Cumulative Quarterly Unedited Data Files 19
   - 2.3 Preliminary Editing and Coding of Processed Interview Data 20
     - 2.3.1 Coding of "OTHER, Specify" Data 20
     - 2.3.2 Investigation of Response Patterns in Usable Records 27
     - 2.3.3 Edits to "Date-Dependent" Variables 28
   - 2.4 General Issues and Principles for Editing NSDUH Data 30
     - 2.4.1 Editing across Modules 30
     - 2.4.2 Assignment of Relevant "Not Applicable" Codes 31
     - 2.4.3 Additional Assignment of NSDUH Missing Data Codes as Part of Editing within a Module 33
     - 2.4.4 Assignment of Codes to Edited "Enter All That Apply" Variables 34
     - 2.4.5 "Back-Editing" Based on "OTHER, Specify" Data 35
     - 2.4.6 "Flag and Impute" Principles for Drug Use Data 35

3. Imputation and the Predictive Mean Neighborhood Methodology 39
   - 3.1 Introduction 39
   - 3.2 Development of the Predictive Mean Neighborhood Methodology 40
     - 3.2.1 Previously Used Hot-Deck Imputation Methods 41
     - 3.2.2 Advantages of the Predictive Mean Neighborhood Methodology 44
   - 3.3 Implementation for the Predictive Mean Neighborhood Methodology 45
     - 3.3.1 Step 1: Response Propensity Adjustment 45
     - 3.3.2 Step 2: Prediction Modeling 46
     - 3.3.3 Step 3: Hot-Deck Imputation 49
   - 3.4 Predictive Mean Neighborhood "Types" 54
     - 3.4.1 Type 1: Single Response Propensity/Single Prediction 55
     - 3.4.2 Type 2: Multiple Response Propensity/Multiple Prediction 56
     - 3.4.3 Type 3: Single Response Propensity/Multiple Prediction 58

Table of Contents (continued)

Chapter **Page**

3.5 Special Auxiliary Variables: Age Group and State Rank 58
3.5.1 Age Groups 59
3.5.2 State Rank 59

4. Editing and Imputation for the NSDUH Core Demographics Variables 61
4.1 Introduction 61
4.2 Editing Selected Core Demographics Variables 62
4.2.1 Creating the Edited Interview Date Variable (INTDATE) 62
4.2.2 Creating the Edited Age Variable (AGE) 63
4.2.3 Creating the Edited Current State Residency Variable (STATELOC) 66
4.2.4 Creating the Edited Gender Variable (IRSEX) 67
4.2.5 Creating the Edited Birth Date Variable (BRTHDATE) 67
4.2.6 Creating the Edited Marital Status Variables (MARITAL, EDMARIT, and NOMARR) 67
4.2.7 Creating the Edited Race and Hispanic/Latino (Origin and Group) Variables 68
4.2.8 Creating the Edited Military Service History and Current Service Variables (SERVICE and MILSTAT) 79
4.2.9 Creating the Edited Highest Grade Completed Variables (EDUC and EDEDUC) 80
4.2.10 Creating the Perceived Health Status Variable (HEALTH) 80
4.3 Imputation for Selected Core Demographics Variables 81
4.3.1 Marital Status Variable (Imputation Set 1) 81
4.3.2 Race Variables (Imputation Set 2) 82
4.3.3 Hispanic/Latino Origin Variable (Imputation Set 3) 85
4.3.4 Hispanic/Latino Group Variable (Imputation Set 4) 86
4.3.5 Education Level Variable (Imputation Set 5) 87

5. Editing and Imputation for the NSDUH Noncore Demographics Variables 91
5.1 Introduction 91
5.2 Editing the Noncore Demographics Variables 91
5.2.1 Moves in the Past 12 Months and State Residency in the Past 2 Years 91
5.2.2 Immigrant Status 94
5.2.3 Noncore Education 95
5.2.4 Employment and Workplace 105
5.2.5 Field Interviewer Debriefing Questions 109
5.3 Imputation for Noncore Demographics Variables 109
5.3.1 Employment Status (Imputation Set 1) 110
5.3.2 Immigrant Status, Born-in-U.S. Indicator (Imputation Set 2) 111
5.3.3 Immigrant Status, Age of Entry (Imputation Set 3) 111

Table of Contents (continued)

Chapter **Page**

6. Editing and Imputation for the NSDUH Core Drug Use Variables 113
   6.1 Introduction 113
      6.1.1 Edited (but Not Imputed) Drug Use Variables 113
      6.1.2 "Parent" and "Child" Drug Use Categories 114
      6.1.3 Special Situation for Methamphetamine, Stimulants, and Psychotherapeutics 115
   6.2 Editing Procedures for the Core Drug Use Variables 116
      6.2.1 Edits of Lifetime Use Variables 117
      6.2.2 Edits of Recency-of-Use Variables 123
      6.2.3 Edits Involving Users of Only Child Drugs 133
      6.2.4 Edits of Frequency-of-Use Variables 135
      6.2.5 Incidence (Age at First Use, Month at First Use, and Year at First Use) 140
      6.2.6 Editing of Noncore Methamphetamine Use Data 142
      6.2.7 Editing of Lifetime Daily Cigarette Use Status 149
      6.2.8 Additional Edits Applied During Imputation Processing 149
      6.2.9 Edits for Drug Variables that Do Not Undergo Imputation 152
   6.3 Imputation of the Core Drug Use Variables 158
      6.3.1 Lifetime Drug Use (Imputation Set 1) 159
      6.3.2 Imputation-Revised Cigarette Recency and Frequency of Use (Imputation Set 2) 162
      6.3.3 Imputation-Revised Cigarette Age at First Use (Imputation Set 3) 166
      6.3.4 Imputation-Revised Age at First Daily Cigarette Use (Imputation Sets 4 and 5) 171
      6.3.5 "Other" Drugs Recency and Frequency 173
      6.3.6 "Other" Drugs Age at First Use 183
      6.3.7 Special Section: Core-Plus-Noncore Methamphetamine and Stimulants Lifetime Use and Recency of Use (Imputation Sets 31 and 32) 186

7. Editing and Imputation for Variables from the NSDUH Noncore ACASI Modules 191
   7.1 Introduction 191
   7.2 Editing of Noncore ACASI Data Because of Patterned Responses 192
   7.3 "Lockout" Following Completion of ACASI Modules 192
   7.4 Editing and Imputation Procedures for Specific Noncore ACASI Modules 193
      7.4.1 Special Drugs Module 193
      7.4.2 Risk/Availability Module 204
      7.4.3 Blunts Module 204
      7.4.4 Substance Dependence and Abuse Module 207
      7.4.5 Special Topics Module 215
      7.4.6 Marijuana Purchases Module 218
      7.4.7 Prior Substance Use Module 220
      7.4.8 Substance Treatment Module 234

Table of Contents (continued)

Chapter **Page**

7.4.9 Health Care Module 245
7.4.10 Adult Mental Health Service Utilization Module 248
7.4.11 Social Environment Module 253
7.4.12 Parenting Experiences Module 253
7.4.13 Youth Experiences Module 260
7.4.14 Mental Health Module 261
7.4.15 Adult and Adolescent Depression Modules 262
7.4.16 Youth Mental Health Service Utilization Module 266
7.4.17 Consumption of Alcohol Module 271

8. Editing and Imputation for the NSDUH Roster Variables 277
8.1 Introduction 277
8.2 Editing the Respondent-Level Detailed Roster Variables 280
8.2.1 Roster Consistency Checks 280
8.2.2 Roster-Level Dataset 282
8.2.3 Roster Edits Involving the Self 282
8.2.4 Roster Edits for Other Household Members 283
8.2.5 Final Edited Respondent-Level Detailed Roster Variables 291
8.3 Editing and Imputation Procedures for Roster-Derived Household Composition Variables 291
8.3.1 Creation of Edited Roster-Derived Household Composition Variables 291
8.3.2 Imputation for Roster-Derived Household Composition Variables 293
8.4 Editing the Proxy Variables 294
8.4.1 Edited Indicator of Potential Proxies in Household (EDFAM18) 295
8.4.2 Editing the Proxy Variables when EDFAM18 = 1 296
8.4.3 Editing the Proxy Variables when EDFAM18 = 0 296
8.4.4 Additional Editing for PRXRELAT 297

9. Editing and Imputation for the NSDUH Income Variables 299
9.1 Introduction 299
9.2 Editing for Binary and Finer Category Income Variables 300
9.2.1 Binary Income Variables 300
9.2.2 Finer Category Income Variables 301
9.2.3 Additional Edits to Income Variables for Inconsistencies or Logical Inferences 302
9.3 Imputation for the Income Variables 302
9.3.1 Imputation for Binary Income Variables (Imputation Set 1) 303
9.3.2 Imputation for Finer Category Income Variables (Imputation Set 2) 308

10. Editing and Imputation for the NSDUH Health Insurance Variables 311
10.1 Introduction 311
10.2 Editing the Health Insurance Variables 312
10.2.1 Editing Because of Incorrect State Information 314

Table of Contents (continued)

Chapter **Page**

10.2.2 Determination of Any Health Insurance Coverage....314
10.2.3 Editing of Health Insurance Follow-Up Variables Based on Current Coverage....316
10.2.4 Miscellaneous Issues for Editing the Health Insurance Variables....316
10.2.5 Editing Any Health Insurance and Private Health Insurance Variables, Imputation Set 1 (Old Method)....318
10.2.6 Editing Types of Health Insurance and Any Other Health Insurance Variables, Imputation Sets 2 and 3 (Constituent Variables Method)....319
10.3 Imputation for the Health Insurance Variables....319
10.3.1 Any Health Insurance and Private Health Insurance Variables, Imputation Set 1 (Old Method)....320
10.3.2 Types of Health Insurance Variables, Imputation Set 2 (Constituent Variables Method)....322
10.3.3 Any Other Health Insurance Variable, Imputation Set 3 (Constituent Variables Method)....324

11. Editing and Imputation for the NSDUH Pair Variables....327
11.1 Introduction....327
11.1.1 Pair Relationship Variable....327
11.1.2 Multiplicity Count Variables....328
11.1.3 Household-Level Person Count Variables....329
11.1.4 Staged Variable Processing....330
11.2 Stage One Editing: Pair Relationships....331
11.2.1 Editing the Household Roster of Each Pair Member....331
11.2.2 Creating the Pair Relationship Variable (PAIRREL)....331
11.3 Stage Two Editing: Multiplicity Counts....337
11.3.1 Determining the Multiplicity Count for Each Pair Member....338
11.3.2 Determining the Final Multiplicity Count....338
11.4 Stage Three Editing: Household-Level Person Counts....339
11.4.1 Determining the Household-Level Person Count for Each Respondent....341
11.4.2 Determining the Final Household-Level Person Count....345
11.5 Imputation for the Pair Variables....346
11.5.1 Creation of Covariates for Imputing Pair-Level Variables....346
11.5.2 Stage One Imputation: Pair Relationships....349
11.5.3 Stage Two Imputation: Multiplicity Counts....352
11.5.4 Stage Three Imputation: Household-Level Person Counts....354

References....359

Table of Contents (continued)

Appendix **Page**

A Percentage of Cases Imputed or Logically Assigned, 2012 Survey A-1

B Detailed Documentation of Edits for Variables from the Noncore Demographics, Core Substance Use, and Noncore Self-Administered Sections in the 2012 NSDUH B-1

C Race and Hispanic/Latino Group Alpha Codes C-1

D Model Summaries D-1

E Hot-Deck Procedure Summaries E-1

F Quality Control Measures Used in the Imputation Procedures F-1

G Interviewer Explanations for Overrides to Consistency Checks in Household Roster G-1

H Rules for Determining Pair Relationships H-1

I Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members I-1

J Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members J-1

K Priority Conditions for Creating Household-Consistent Covariates K-1

L Creation of Household-Level and Person-Level Files L-1

List of Tables

Table		Page
1.1	Content of Core Modules in the 2012 NSDUH Interview	3
1.2	Content of Noncore ACASI Modules in the NSDUH Interview	4
2.1	Effects on the Potential Usable Status of a Case Based on Responses to Follow-Up Questions for Refusals to Gate Questions in the Core Drug Modules	18
3.1	Summary of Item Imputation Procedure Used, by Variable and NSDUH Year	40
3.2	Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood	47
3.3	Values of Delta for Various Predicted Probabilities	51
3.4	Examples of Variables Imputed Using Each of the Four Combinations of Univariate/Multivariate Matching and Assignment	53
3.5	PMN Types Applied to Each Variable Group and Imputation Set	55
4.1	Hispanic/Latino Status of Item Nonrespondents for Race	84
4.2	Edited Race Variables and Their Imputation-Revised Counterparts	84
4.3	Mapping of EDHOGRP Levels to EDHOGRP2 Levels	87
4.4	Mapping of EDEDUC Levels to RESPEDUC Levels	88
5.1	Mapping of Noncore Education Questions to Edited Variables	97
5.2	Mapping of Current Age with Expected Grades	101
5.3	Categories of JBSTATR	107
6.1	Core Drugs in the 2012 NSDUH with a Parent/Child Relationship	115
6.2	How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables	127
6.3	How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables	131
6.4	Mapping of Noncore Methamphetamine Questions in the Special Drugs Module to Edited Variables	143
6.5	Drugs and Drug Use Measures, Imputation Sets	158
6.6	General Incomplete Recency Categories for Cigarettes	164
6.7	Elements of Full Predictive Mean Vector: Cigarettes	166
6.8	Elements of Full Predictive Mean Vector	182
6.9	Prevalence Recodes Incorporating More than One Recency Variable	183
7.1	Mapping of Cigarette Dependence Questions to Edited Variables	212

List of Tables (continued)

Table **Page**

7.2 Summary of Response Patterns for 2012 NDSS Variables 215
7.3 Depression Symptom and Score Variables.....265
8.1 Roster Grid Example Where Number of Persons in Household (QD54) Equals 4 277
8.2 Roster Relationship Codes.....278
8.3 Roster-Derived Household Composition Variables.....278
8.4 Proxy Variables.....279
8.5 Levels of PRXRELAT.....295
9.1 Comparison between 1999-2007 and 2008-2012 Sets of Income Questions 299
9.2 Imputation Order for Binary Income Variables (Imputation Set 1) 303
9.3 Cross-Classification of FAMPMT and FAMSVC; Relevant Combinations for Binary Income Predictive Mean Vector when WELMOS Is Missing.....307
10.1 Edits Pertaining to the Health Insurance Section.....317
10.2 Mapping of Raw Health Insurance Variables to Base Variables for Imputation Set 1 (Old Method) 318
11.1 Edited and Imputation-Revised Multiplicity Count Variables 329
11.2 Edited and Imputation-Revised Household-Level Person Count Variables.....330
11.3 Levels of the Variable PAIRREL 332
11.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH.....335
11.5 Frequencies of the Levels of the Variable RELMATCH: 2012 336
11.6 Multiplicity Counts for Each Pair Member 338
11.7 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2012.....339
11.8 Age Group Pairs with Associated Possible Pair Relationships 349
11.9 Modeled Pair Relationships within Age Group Pairs.....351
11.10 Modeled Multiplicities within Domains.....353

List of Figures

Figure **Page**

3.1 Donor Selection Algorithm 50

3.2 PMN Type 1: Single Response Propensity/Single Prediction 55

3.3 PMN Type 2: Multiple Response Propensity/Multiple Prediction 57

3.4 PMN Type 3: Single Response Propensity/Multiple Prediction 58

5.1 Follow-Up Question Logic for Respondents in the Noncore Education Module 96

5.2 Routing Logic for Respondents in the Employment Status Module 106

6.1 Routing Logic for Respondents in the Core Drug Modules 114

6.2 Routing Logic and Editing Overview for Methamphetamine and Other Stimulant Questions in the Special Drugs Module 144

10.1 Routing Logic for Respondents in the Health Insurance Module 313

1. Introduction

Conducted annually, the National Survey on Drug Use and Health (NSDUH) is the primary source of information on the prevalence, patterns, and consequences of alcohol, tobacco, and illegal drug use and abuse among all U.S. civilian, noninstitutionalized residents of the 50 States and the District of Columbia, aged 12 or older. In the 2012 NSDUH, this population included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes) and civilians residing on military bases. The target population excluded persons with no fixed household address (e.g., homeless transients not in shelters), residents of institutional group quarters (e.g., jails and hospitals), children younger than 12, and active military personnel. As it has since 1999, the 2012 NSDUH utilized a 50-State, multistage cluster design that enables the Substance Abuse and Mental Health Services Administration (SAMHSA) to provide representative estimates for each State and the District of Columbia. Both direct and model-based State and substate estimates are produced on a variety of measures based on a combination of multiple years of data.

This report focuses on the editing and statistical imputation procedures that were used with respondent data for the 2012 NSDUH. Logical editing uses data from elsewhere within the same respondent's record to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables. Imputation is defined as the replacement of missing values with valid, nonmissing values. Statistical imputation usually involves some randomness to preserve the natural variability in the data.

1.1 Organization of the NSDUH Questionnaire and Overview of Content

The 1999 survey marked the transition from data collection based on paper-and-pencil interviewing (PAPI) to computer-assisted interviewing (CAI).[1] Since then, the NSDUH data have been collected using CAI, which allows a private mode of data collection for respondents to answer questions pertaining to drug use and other sensitive topics.

The CAI interview is organized by modules. "Modules" in the NSDUH questionnaire refer to sections that are organized together by mode of administration, content, and routing logic. For example, the alcohol module includes an initial question that asks whether respondents ever had a drink of an alcoholic beverage. Respondents who report that they have ever used alcohol are asked additional questions about their age when they first used alcohol, when they last used alcohol, and (if relevant) their use of alcohol in the 12 months or 30 days prior to the interview. Respondents who do not report lifetime alcohol use are routed to the next module, which pertains to marijuana use.

The first module consists of questions about certain demographic characteristics, including birth date (which is used to determine a respondent's age), gender, marital status, Hispanic/Latino origin, racial group, and education level (highest grade completed). Computer-assisted personal interviewing (CAPI) is used for these questions, in which interviewers read the

[1] CAI specifications for the 2012 NSDUH are available at http://www.samhsa.gov/data/2k12/NSDUH2012MRB/2k12Q.pdf.

questions and respondents give their answers aloud to the interviewers, who then enter the responses into the computer. The logic for determining which questions the interviewers should ask is controlled by the computer program based on the responses to previous questions that interviewers enter into the computer. Consequently, interviewers can concentrate on asking questions and recording respondent answers, without having to concern themselves with comprehending and following skip pattern instructions.

Following completion of these demographic questions, interviewers orient respondents about the use of specific features of the laptop computer and use of headphones for listening to questions. Respondents then complete a brief tutorial on answering self-administered questions, which are administered through use of audio computer-assisted self-interviewing (ACASI), ACASI is used for the majority of questions in the interview and is particularly useful for collecting data about sensitive topics such as substance use, problems associated with substance use, risk and protective factors related to substance use and mental health issues.

In the ACASI portion of the interview, respondents can read the questions on a computer screen and also are encouraged to listen to an audio recording of the questions on headphones. Respondents then enter their answers directly into a laptop computer. These features of ACASI prevent interviewers or others in the household from knowing what questions the respondents are being asked and how they are answering. The availability of audio recordings of the questions is especially useful for respondents with limited reading ability because they can listen to the questions instead of having to read them.

Once respondents complete the ACASI portion of the interview, they turn the laptop computer back over to the interviewer. Remaining interview questions are administered through CAPI. Topics in the remainder of the interview include immigrant status (i.e., whether respondents were born in the United States), State residency in periods prior to the interview, current education status, employment status and workplace issues, household characteristics, health insurance coverage, and income.

In addition, the CAI instrument consists of core and noncore modules. Core modules, such as those pertaining to key demographic characteristics and substance use, have been designed to stay relatively constant from one year to the next to permit measurement of trends in drug use across time. Table 1.1 summarizes the content of core modules in the 2012 NSDUH interview, including the mode of administration. For prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), the questionnaire asked about "nonmedical" use. Nonmedical use is defined in NSDUH as use of a drug without a prescription of the respondent's own or use only for the experience or feeling that the drug caused. In addition, for the questions for lifetime nonmedical use of prescription drugs, respondents were shown printed "pill cards" with pictures of prescription drugs for questions in the module to aid respondents in answering the questions.

Table 1.1 Content of Core Modules in the 2012 NSDUH Interview

Module	Content	Mode of Administration
Core Demographics	• Age • Gender • Hispanic/Latino origin and race • Marital status • Military service • Highest educational grade • Perceived health status	Interviewer administration using CAPI
Tobacco	• Lifetime use or nonuse of the following: – Cigarettes – Snuff – Chewing tobacco – Cigars – Pipe tobacco • Additional questions for lifetime users	Self-administration using ACASI
Alcohol	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Marijuana	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Cocaine and Crack	• Lifetime use or nonuse of the following: – Any cocaine – Crack cocaine (if lifetime cocaine user) • Additional questions for lifetime users	ACASI
Heroin	• Lifetime use or nonuse • Additional questions for lifetime users	ACASI
Hallucinogens	• Lifetime use or nonuse of seven hallucinogens, including any other hallucinogen besides the ones that had been listed • Additional questions if lifetime use reported for any of the seven specific hallucinogens	ACASI
Inhalants	• Lifetime use or nonuse of 11 specific types of inhalants for kicks or to get high, including any other inhalant besides the ones that had been listed • Additional questions if lifetime use reported for any of the 11 specific inhalants	ACASI
Pain relievers	• Lifetime use or nonuse of any of 28 specific prescription pain relievers that were not prescribed or were taken only for the experience or feeling (i.e., "nonmedical" use), including any other prescription pain reliever besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific pain relievers	ACASI, plus a printed "pill card" showing pictures of specific pain relievers to aid respondent recall
Tranquilizers	• Lifetime nonmedical use or no nonmedical use of any of 22 specific prescription tranquilizers, including any other prescription tranquilizer besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific tranquilizers	ACASI, plus a printed pill card showing pictures of specific tranquilizers to aid respondent recall

Table 1.1 Content of Core Modules in the 2012 NSDUH Interview (continued)

Module	Content	Mode of Administration
Stimulants	• Lifetime nonmedical use or no nonmedical use of any of 19 specific prescription stimulants or types of stimulants (e.g., prescription diet pills), including methamphetamine (which often is illegally manufactured) and any other prescription stimulant besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific stimulants or types	ACASI, plus a printed pill card showing pictures of specific stimulants to aid respondent recall
Sedatives	• Lifetime nonmedical use or no nonmedical use of any of 15 specific prescription sedatives or types of stimulants (e.g., barbiturates), including any other prescription sedative besides the ones shown to the respondent • Additional questions if lifetime nonmedical use reported for any of the specific sedatives or types	ACASI, plus a printed pill card showing pictures of specific sedatives to aid respondent recall

ACASI = audio computer-assisted self-interviewing; CAPI = computer-assisted personal interviewing.

In contrast to the core modules, the content of noncore modules can change across years to measure new or developing topics of interest or to rotate certain topics in or out of the interview. In noncore sections, therefore, questions or entire modules can be added or deleted, or the wording of existing questions can change from one year to the next. The topics that are covered in noncore modules also can show more variation than the topics that are included for the core modules. As shown in Table 1.2, for example, there were 18 noncore ACASI modules in 2012 that covered topics such as injection drug use, perceptions of risk and availability for different substances, substance use disorders (i.e., dependence or abuse), arrests and driving under the influence of alcohol or illicit drugs, treatment for substance use problems, physical health conditions, risk and protective factors for substance use among adolescents, and mental health issues.

Table 1.2 Content of Noncore ACASI Modules in the NSDUH Interview

Module	Content
Special Drugs	• Routes of heroin administration (i.e., smoking, sniffing, injection) • Injection of methamphetamine, other stimulants, cocaine, or other drugs • Additional methamphetamine questions • General injection use behaviors • Miscellaneous drugs (e.g., Adderall®, Ambien®, cough and cold medications)
Risk/Availability	• Perceived risk of harm associated with use of cigarettes, alcohol, or specific illicit drugs • Perceived availability of illicit drugs • Miscellaneous risk behaviors (e.g., seatbelt use)
Blunts	• Use of cigars with marijuana in them (i.e., "blunts")
Substance Dependence and Abuse	• Nicotine dependence • Symptoms of dependence and abuse for alcohol and illicit drugs, including nonmedical use of prescription drugs
Special Topics	• Arrests in the lifetime and past 12 months • Probation and parole status, past 12 months • Driving under the influence of alcohol or illegal drugs • Knowledge of State laws for marijuana possession

Table 1.2 Content of Noncore ACASI Modules in the NSDUH Interview (continued)

Module	Content
Market Information for Marijuana	• How respondents who used marijuana in the past 12 months obtained marijuana, including price/value information for marijuana that was purchased or obtained through a trade
Prior Substance Use	• Use of marijuana, cigarettes, alcohol, or cocaine more than 12 months ago but within the past 24 months • Last use of substances included in the core drug section (see Table 1.1) if not used in the past 30 days • Sources of prescription drugs and methamphetamine • Sequence of initiation of cigarettes, alcohol, and marijuana for adolescents aged 12 to 17
Substance Treatment	• Treatment for use of alcohol or illicit drugs (i.e., not counting tobacco) in the lifetime and past 12 months • Perceived need for treatment for use of alcohol or illicit drugs • Barriers to receiving treatment for use of alcohol or illicit drugs
Health Care	• Pregnancy status of females aged 12 to 44 • Emergency room visits and hospitalizations in the past 12 months • History of specific medical conditions
Adult Mental Health Service Utilization	• Administered to respondents aged 18 or older • Use of inpatient or outpatient mental health services in the past 12 months • Payment for inpatient or outpatient mental health services • Use of prescription medication to treat a mental health condition • Unmet demand for mental health services • Use of alternative sources for mental health treatment
Social Environment	• Administered to respondents aged 18 or older • Moves in the past 5 years • Specific illegal behaviors regardless of arrest (e.g., selling illegal drugs) • Attitudes about marijuana use • Religious involvement and beliefs
Parenting Experiences	• Administered to parents when two persons were selected for an interview and the second selected person was an adolescent • Beliefs about whether their child has used tobacco, alcohol, or drugs • Talks with their child about substance use • Attitudes about drug education
Youth Experiences	• Administered to respondents aged 12 to 17 • Moves in the past 5 years • Risk and protective factors for substance use • Fighting and delinquent behaviors • Attitudes and perceptions of others' attitudes about substance use • Exposure to drug prevention messages • Religious involvement and beliefs
Mental Health	• Administered to respondents aged 18 or older • Psychological distress in the past 30 days or past 12 months • Impairment in carrying out activities because of psychological distress • Suicidal thoughts and behavior
Adult Depression	• Administered to respondents aged 18 or older • Symptoms of depression in the lifetime and past 12 months • Impairment in carrying out activities because of depression symptoms • Treatment for depression

Table 1.2 Content of Noncore ACASI Modules in the NSDUH Interview (continued)

Module	Content
Youth Mental Health Service Utilization	• Administered to respondents aged 12 to 17 • Use of mental health services in the past 12 months • Reasons for receiving mental health services from specific sources
Adolescent Depression	• Administered to respondents aged 12 to 17 • Age-appropriate questions for symptoms of depression in the lifetime and past 12 months • Age-appropriate questions for impairment in carrying out activities because of depression symptoms • Treatment for depression
Consumption of Alcohol	• Number of drinks on the last occasion of alcohol use in the past 30 days • Underage alcohol use • Alcohol use in combination with illicit drugs • Initiation of consumption of five or more drinks on an occasion (i.e., binge alcohol use) • Females' consumption of four or more drinks on an occasion

ACASI = audio computer-assisted self-interviewing.

1.2 Overall Data Quality Issues with CAI

Conversion of the NSDUH interview from a paper-and-pencil format (PAPI) to a computer-assisted format (CAI) greatly reduced or in some cases eliminated the following data quality problems:

- illegible responses, multiple marks, or out-of-range values;
- item nonresponse (i.e., missing data);
- incorrectly executed skip patterns; and
- inconsistencies among related variables.

For example, when a question's instructions ask respondents to choose only one answer, PAPI respondents nevertheless are physically able to mark multiple answers. This cannot occur in the CAI because the computer program will permit entry of only one response to the item. Similarly, the CAI has been programmed not to allow out-of-range values for certain items, such as frequency-of-use items, thereby reducing the amount and types of out-of-range values that would otherwise need to be addressed through editing. Further, the skip patterns that are embedded in the CAI were designed to reduce the occurrence of inconsistent data by not giving respondents the *opportunity* to provide inconsistent answers. The occurrence of inconsistent data was further reduced through the use of consistency checks built into the CAI program that prompted respondents to resolve inconsistencies that occurred between related items.

Despite the potential for improvements in data quality through a CAI instrument, it was recognized that conversion to CAI would not completely eliminate data problems. For example, missing data were not completely eliminated because CAI respondents still had the option of entering a response of "don't know" or "refused" when answering a given item. Similarly, even though consistency checks were designed to reduce inconsistent reporting, the CAI program was not equipped to address every possible inconsistent report that a respondent could make. Finally,

in some situations, conversion to CAI could introduce new data quality issues. As discussed in Section 2.3.2 in Chapter 2, for example, the lack of direct interviewer monitoring could allow some respondents to use the computer keyboard to enter nonsensical patterns of answers for reasons such as accidental errors or disinterest.

1.3 Organization of this Report

The material in this report combines two sections from previous versions of the NSDUH Methodological Resource Book (MRB): Section 10, which addressed logical editing and coding of variables, and Section 11, which addressed the statistical imputation procedures that were applied to variables that underwent additional processing. These two sections have been combined for the 2012 MRB for several reasons.

First, editing and imputation are closely related, and combining the reports affords the opportunity to remove redundant information.[2] Second, structuring the report in this manner is designed to aid readers in following the "life cycle" of NSDUH data, starting with a respondent's answers to NSDUH questions, how the variables that capture these answers are edited, and, where applicable, how the data are imputed after having been edited. Third, in MRBs prior to 2012, some documentation of editing procedures appeared in Section 11 instead of in Section 10; thus, readers who were primarily interested in documentation of the editing procedures that were applied to certain variables had to check two MRB sections to locate that information rather than just one.

Preliminary coding and processing of unedited data after interviewers transmitted the data from the field as well as the general principles of logical editing are discussed in Chapter 2 of this report. The predictive mean neighborhood (PMN) imputation methodology, which is used to impute NSDUH data, is described in detail in Chapter 3. It is recommended that readers first review Chapters 2 and 3, as these two chapters provide a foundation for Chapters 4 through 11, The information in both chapters helps to set an appropriate context for readers as they review the documentation for the specific editing and imputation procedures that were employed on a particular variable set.

Following these initial chapters, Chapters 4 and 5 address the editing and imputation procedures that were applied to the CAI core and noncore demographic variables, respectively. Editing and imputation procedures for the core substance use variables are discussed in Chapter 6, and Chapter 7 addresses the editing procedures for the noncore variables. Chapter 7 also discusses the imputation procedures for cigarette dependence,[3] which differ from the procedures used for other drug variables. Chapter 8 describes the edits applied to the household roster, the creation of imputation-revised versions of the roster-derived household composition variables, and the creation of respondent-level variables with individual roster information. Chapter 9 summarizes the editing and imputation procedures that were applied to the income variables. Procedures for logical editing and imputation of missing values in the health insurance variables

[2] The similar themes running through Appendices B and E of this report provide a good illustration of the close relationship of editing and imputation. In particular, the edits listed in Appendix B and the logical constraints listed in Appendix E overlap substantially.

[3] This term is referred to in past imputation reports as "nicotine dependence."

are described in Chapter 10. The editing and imputation processing of the pair relationship variable and related household count variables are detailed in Chapter 11.

This report also contains 12 appendices. Appendix A of this report is identical to Appendix A of the 2012 Editing and Imputation Evaluation Report (Scott et al., 2013). It contains a number of tables that quantify the amount of imputation and logical assignment (i.e., editing) that selected analytic variables underwent during imputation processing in 2012. Appendix B provides a summary of data issues and the specific edits that were applied in response to these issues for the noncore demographic, core drug use, and noncore drug use variables. Appendix C provides details on the handling of "OTHER, Specify" responses to the race and Hispanic/Latino group questions so that the data could be summarized in a meaningful way. The covariates in each of the imputation models are listed in Appendix D. The tables in Appendix D also include (1) the starting list of covariates for each model and (2) descriptions of each level and identification of the reference level for categorical covariates. Appendix E provides details on each final hot-deck step in the imputation procedures. The quality control measures used in the imputation procedures are summarized in Appendix F. Reasons that interviewers gave for overriding consistency checks in the household roster are presented in Appendix G, along with evaluations of their legitimacy and the resulting actions in editing the roster. The rules for determining pair relationships are defined in Appendix H. The conditions used for reconciling differing multiplicity counts between pair members are described in Appendix I, and the conditions used for reconciling differing household-level person counts between pair members are described in Appendix J. Appendix K details the priority conditions for creating household-consistent covariates. Appendix L contains detailed information about household-level and person-level eligibility and the completeness criteria used to construct the household-level and person-level files.

1.4 Changes from the 2011 Survey to the 2012 Survey

This section summarizes the changes in editing and imputation procedures implemented on the NSDUH since the 2011 survey year. It is intended for analysts who have had previous experience with the data. Those with little or no prior experience with NSDUH data are encouraged to review Chapters 2 and 3 before reviewing this section. Before making a change to the editing and imputation process, a careful assessment of the impact of the change is done to determine what, if any, effect the change would have on the estimates. The impact assessment involves applying the new process to the previous year's data and comparing the results. For variables whose creation has a stochastic component, the new process is run at least three times to assess whether the results are directly attributable to the new process or whether they could be due to random variation.

Overall, there were no appreciable changes to the editing procedures and only minor changes to the imputation procedures. The changes to the imputation procedures are described in three sections below. The first section describes changes that were implemented to handle atypical cases appearing in the 2012 data. The second section describes two modifications that reduced the time required to process the data, with little impact on the final imputation-revised variables. Finally, the third section lists a few minor corrections and improvements that also had little impact on the final imputation-revised variables. The changes described in these three sections were implemented during the normal imputation process.

1.4.1 Changes to Accommodate Atypical Cases

In almost every year of the NSDUH, a few cases arise with patterns of response that have not been seen in recent years. Often, the editing and imputation procedures have to be modified slightly to accommodate these cases. In 2012, there were three changes to the imputation procedures that were implemented when unusual cases were encountered. These changes are described below.

The first unusual case in the 2012 data involved a respondent with a 12-month crack frequency of 364 days. This value implied that the 12-month frequency for cocaine was either 364 or 365 days. However, no donor was available with a 12-month cocaine frequency as large as 364. For this individual, a random stochastic assignment was made to ensure the 12-month frequency for cocaine was 364 or 365 days.

The second unusual instance to arise in 2012 involved the imputation of heroin recency and frequency. During the process of modeling the probability of past month heroin use, it was discovered that no respondents in the 12-17 age group reported past month use of heroin. Because no respondents were available to build the model, it was determined that the best course of action would be to assign each nonrespondent in this age group a predicted mean of zero for past month heroin use. The alternative would be to leave the probability of past month use as missing, but this could have caused problems when calculating the Mahalanobis distance in the final donor selection step.

Finally, in the 2012 data, a respondent did not respond affirmatively to the question about Hispanic/Latino origin, but listed "Puerto Rican" as his or her only race. When Hispanic/Latino respondents list "Puerto Rican" as their only race, likeness constraints require the donor for the race variable to be Hispanic/Latino and also to list "Puerto Rican" as one of their Hispanic/Latino groups. A new missingness pattern was set up to handle this specific respondent. This new missingness pattern did not include a likeness constraint requiring the donor to be Hispanic/Latino, but it did include a likeness constraint requiring the donor to be Puerto Rican. The donor could be classified as Puerto Rican either by selecting "Puerto Rican" as a Hispanic/Latino group or by entering "Puerto Rican" as a write-in response.

1.4.2 Changes to Streamline Imputation Procedures

In 2012, two measures to streamline the imputation process were undertaken and are described below.

During 2012 processing, the imputation procedures for the six race indicator variables IRRACE*XX* and the multiple Hispanic/Latino group variable IRHOGRPM were streamlined to improve efficiency. Prior to 2012, the six IRRACE*XX* variables were assigned values using the same donor that was used in the hot-deck step for the race variables IRDETAILEDRACE and IRNWRACE but in a separate program. Similarly, IRHOGRPM was assigned values using the same donor that was used for the Hispanic/Latino group variable IRHOGRP4 but in a separate program. Beginning with the 12-month processing in 2012, the later programs were eliminated and the sets of variables that underwent imputation in the earlier programs were augmented. The elimination of the two programs allowed for more efficient processing of the demographic

variables. The earlier structure existed because the decisions to create the IRRACE*XX* variables and the IRHOGRPM variable were made after the original program structure was developed.

Another change was made to the demographic processing procedures in 2012. Prior to 2012, a variable called IRDETAILEDRACE was created for use as a likeness constraint for the Hispanic/Latino group. The creation of this variable was eliminated in 2012, and the constraint was rewritten using the six IRRACE*XX* variables. This eliminated the need to create IRDETAILEDRACE.

1.4.3 Changes Involving Minor Corrections and Recodes

Seven changes were made to the 2012 imputation procedures that involved minor recodes or corrections. None of these had a significant impact on the estimates.

The first correction involved a case in the 2012 data where the respondent reported past year but not past month use of pain relievers and lifetime use or nonuse of OxyContin® was unknown. The individual was subsequently imputed to lifetime use of OxyContin® and no lifetime use of other pain relievers. In this case, the individual should have had his or her OxyContin® recency set equal to overall pain reliever recency. However, a small coding error was found that allowed the individual to be imputed to lifetime not past year use of OxyContin®. This code was corrected during 2012 processing and comparisons were made to ensure that no other data were affected. Because of the stringent quality control checks in place, it is unlikely that this correction would affect previous years of data because the type of inconsistency that would result (i.e., the "parent" drug having a more recent value than a "child" drug) is checked multiple times throughout the process, and therefore, similar to 2012, it would have been detected.

The second correction was made during 2012 processing when a small error was found that affects the imputation indicators IIALCD5 and II2ALCD5 (both are associated with past month binge alcohol use). The variable ALDYSFG is an indicator of revision to the edited alcohol 30-day frequency variable ALCDAYS. If a respondent reports a binge drinking 30-day frequency that is greater than the overall alcohol 30-day frequency, he or she is asked which response is correct and is given the opportunity to change the incorrect response. If the responses to the two questions are still inconsistent, ALDYSFG is set to 1, and ALCDAYS is edited to be consistent with the edited binge drinking 30-day frequency variable DR5DAY.

Prior to 2012, the imputation procedures were incorrectly assigning the imputation indicators associated with the binge drinking 30-day frequency. When ALCDYSFG was 1, these indicators, IIALCD5 and II2ALCD5, were being set to 2, indicating logical assignment. Because it is the overall alcohol 30-day frequency that is being logically assigned (not the binge drinking 30-day frequency), these indicators should not have been set to 2; they should have been left as 1 (indicating a valid response). The imputation indicators for the overall alcohol 30-day frequency should have been set to 2, but it turned out that they already were being set to 2 because of other editing flags.

From 2008 to 2011, between 19 and 31 records had IIALCD5 and II2ALCD5 incorrectly assigned a value of 2 instead of 1. This error did not directly affect any other variables, and the

imputation indicators for the binge drinking 30-day frequency-of-use variables are not used in the creation of any other variables. This error was fixed for the 2012 cycle and beyond. Because this error affects so few observations and no other variables were affected other than the imputation indicators (i.e., not the actual imputed value), no changes were made for previous years.

A third change that occurred during 2012 processing involved the imputation indicators for the multiplicity counts. The general principle for all variable groups (drug, demographic, income, insurance, roster, and roster pair variables) is that all respondents who are imputed to be in the domain and all respondents who are imputed to be out of the domain for a certain variable receive a value of 3 (statistically imputed) for the imputation indicator. Prior to 2012 processing, it was discovered that the multiplicity counts did not follow this principle. For variables other than multiplicity counts, respondents who are imputed receive a value of 3 (statistically imputed) for the imputation indicator. The imputation-revised pair relationship variable IRPRREL defines the domains for the multiplicity counts. Some pairs undergo imputation for IRPRREL. The edited version of IRPRREL is called PAIRREL. Several of the levels of PAIRREL call for a restricted imputation; that is, the donor's values of PAIRREL are directly restricted. If some, but not all, of the values of the restricted set are in the domain for the given multiplicity count, then the imputation indicator for the given multiplicity count should indicate that the value for that case was imputed and was not a legitimate skip. Prior to 2012, for the multiplicity counts, respondents who were imputed out of the domain were given imputation indicator values of 9 (legitimate skip), not 3 (statistically imputed). There are also some respondents who get imputed into the domain but do not require imputation for the multiplicity count once that occurs. These respondents were given imputation indicator values of 1, not 3. Because these values are both misleading and inconsistent with what is done for the other variables, this process was changed for 2012 processing. These changes also affected the imputation indicators for some of the household-level person counts because for respondents in some pair domains, household-level person counts are set equal to the opposite-focus multiplicity counts. Because this also affected only the imputation indicator variable and not the actual data, no changes were made to prior years and it did not affect 2012 estimates.

The fourth change involved the hot-deck step for age of first daily cigarette use. Prior to 2012, the last attempt to find a donor involved reimputing the age of first cigarette use for recipients to expand the potential donor pool. This expanded the pool of potential donors by eliminating the need for the donor to have an age of first daily cigarette use greater than or equal to the recipient's imputed age of first cigarette use. Beginning in 2012, instead of reimputing the recipient's age of first cigarette use, the last likeness constraint was eliminated from the last attempt to find a donor. This is a solution that is easier to implement and document, involves shorter run times, and is less prone to error. Moreover, only one or two cases at most are affected by this each year.

Most of the imputation programs use a common SAS program to implement the hot-deck step of the PMN imputation method. This is referred to as the PMN Hot-Deck Common Code. A fifth change involved modifying the number of attempts to find a donor in the PMN Hot-Deck Common Code. Prior to 2012, the number of attempts to find a donor was static for any imputation procedure using the Hot-Deck Common Code. If the missingness patterns varied with respect to the number of tries, the Hot-Deck Common Code would fill in the remaining tries with

empty lists of likeness constraints, essentially dropping all constraints in the process. However, beginning in 2012, the number of attempts to find a donor was modified to be specific to each missingness pattern. This change was made to allow greater control over the hot-deck step and to prevent inadvertently dropping all likeness constraints.

A minor change was also made beginning in 2012 to the level description for PAIRREL=14. The level description was changed from "Pair type not clr, cld be any pr codes 1 thru 13" to "Pair type not clear." This was done because for most of the pairs with PAIRREL=14, the possible values for the imputation-revised version IRPRREL were restricted—not all codes 1 to 13 were possible.

The only other significant change to the procedures in 2012 involved the continuing migration of hot-deck programs to the PMN Hot-Deck Common Code. See Ault et al. (2010) for a description of the reasons for these modifications. More migrations are planned for 2013 processing as time and budget allow.

2. Procedures and General Principles for Editing the Computer-Assisted Data

2.1 Introduction

This chapter provides an overview of the procedures and general principles for editing the computer-assisted interviewing (CAI) data from the National Survey on Drug Use and Health (NSDUH). Logical editing typically uses data from elsewhere within the same respondent's record in a deterministic manner to reduce the occurrence of missing or ambiguous data or to resolve inconsistencies between related variables.[4] In contrast, statistical imputation procedures (see Chapter 3) apply probabilistic (stochastic) statistical methods to identify *another* respondent (i.e., a "donor") whose data are used to replace (1) missing values in the "recipient" respondent's record with nonmissing values from the donor; or (2) ambiguous responses in the recipient respondent's record with more specific information from the donor.

As an example of ambiguous data, the CAI logic requires respondents to report that they have used a particular substance (e.g., marijuana) at least once in their lifetime in order to be asked when they last used it. However, respondents can report lifetime use but not provide a definitive answer for when they last used the substance. In the subsequent imputation procedures for this recipient record, a donor is identified who specifically reported last using the substance in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago. Here, the ambiguous answer in the recipient's record for most recent use is replaced with one of the more specific responses supplied by the donor record.

Section 2.2 describes procedures for initially processing the transmitted NSDUH interview data to get it in a form for further data processing. Subsequent editing and coding steps are described in Section 2.3. The final section in this chapter, Section 2.4, discusses additional principles of data processing and editing that were applied once the transmitted interview data had been processed and cases with questionable data had been identified.

2.2 Initial Processing of Transmitted NSDUH Interview Data

The collected interview data were transmitted from the field as ASCII files, and daily SAS® datasets were created from these files. This daily processing included the following activities as part of the creation of initial unedited interview data files:

[4] There are a few situations where data from outside the respondent's record is used in logical editing. For example, some editing procedures involve data from the screener where an eligible member of the dwelling unit reports basic information about all members of the dwelling unit. Further, in situations where two members of the same dwelling unit are selected for the survey and complete the interview, data from the second respondent's record may be used in logical editing of some variables in the first respondent's record, or vice versa. This procedure allows use of information from both respondents to determine relationships among household members.

- assignment of standard NSDUH missing data codes (Section 2.2.1),
- remapping of responses to "enter all that apply" questions (Section 2.2.2), and
- identification of "usable" cases (Section 2.2.3).

Each day's processed SAS dataset was merged with the transmitted data to date until the end of the quarter (e.g., January through March = Quarter 1) when a cumulative data file (Section 2.2.4) was produced that contained all transmitted cases from that quarter.

2.2.1 Assignment of Standard NSDUH Missing Data Codes

A key activity in the initial processing of transmitted interview data involved the assignment of standard NSDUH missing data codes. The Blaise program for the CAI instrument uses codes of 8 (or 98 or 998, etc.) to denote responses of "refused" and codes of 9 (or 99 or 999, etc.) to denote responses of "don't know." However, in the NSDUH, a code of 98 is used to indicate when a variable is blank (i.e., not answered), and a code of 99 is used in editing procedures to indicate when the variable corresponding to a question was legitimately skipped. Therefore, the codes for missing data that were supplied by the Blaise program were replaced with the standard NSDUH codes for "don't know" (DK) and "refused" (REF). Assignment of codes as part of the editing procedures to indicate when a question was legitimately skipped is discussed in Section 2.4.2.

The following standard codes for missing data were relevant to the 2012 CAI data, depending on the number of digits for a given variable:

- 94 (or 994 or 9994, etc.) = DON'T KNOW (DK);
- 97 (or 997 or 9997, etc.) = REFUSED (REF); and
- 98 (or 998 or 9998, etc.) = BLANK (i.e., nonresponse [NR]).

Codes for missing data in most unedited variables were two digits in length (i.e., 94, 97, or 98). For some variables, however, these values were part of the allowable range of responses. Questions that asked respondents to report the age when they first used a particular drug, for example, had an allowable range of up to 110 years. For the variables corresponding to these age-related questions, the codes for missing data were three digits in length (i.e., 994, 997, or 998).

Finally, the CAI logic governed whether respondents were asked additional questions about a topic based on their age, gender, or answers to other preceding questions. For example, questions in the interview about pregnancy applied only to females aged 12 to 44 (see Section 7.4.9 in Chapter 7). These pregnancy questions were skipped if interviewers reported that a respondent was male or if the CAI program recorded that a female respondent was aged 45 or older. When questions were skipped because the criteria in the CAI program were not satisfied for administering the questions, the unedited variables corresponding to the skipped questions retained a code of "blank" (i.e., 98, 998, etc.). In subsequent editing (described in Section 2.4.2), these variables were examined more closely to determine whether the questions had been legitimately skipped (i.e., they were not applicable) or whether they should retain codes of "blank."

2.2.2 Remapping of Responses to "Enter All that Apply" Questions

A second activity associated with the initial processing of transmitted interview data involved the remapping of responses to "enter all that apply" type questions, which allowed respondents to choose as many responses from a given list as applied. Respondents who wanted to report more than one answer from a list did so by typing the numeric codes that corresponded to the applicable responses and by separating each entry with a space.

The CAI program captured information from these "enter all that apply" questions as separate variables in the order that respondents keyed their answer choices. For example, the CAI program included 18 separate variables for question PR04A to accommodate reports of lifetime nonmedical use of pain relievers from the list. Consequently, an "enter all that apply" variable in the transmitted data could have a different meaning across respondents, depending on which answer a respondent chose first and the number of answers that the respondent chose from the list. For example, if a respondent reported only nonmedical use of the pain reliever OxyContin® from the list in question PR04A, this response would be captured in the *first* variable from the transmitted data. The 17 remaining "unused" variables from question PR04A would be blank. If another respondent chose codeine and morphine as his or her first and second responses from the list and then chose OxyContin®, then the first variable from PR04A would be occupied by the response for codeine, the second variable would be occupied by morphine, the third would be occupied by OxyContin®, and the remaining variables would be blank.

If these "enter all that apply" variables were allowed to remain as in the transmitted data from the CAI, then these variables would have retained the information about the order in which respondents chose their answers. However, this variable structure makes it difficult to analyze the data. For example, it would be more straightforward for information about lifetime nonmedical use of OxyContin® to be captured in a single variable for subsequent use in creating edited and imputed variables for this measure. Otherwise, 18 different variables would need to be checked to identify reports of nonmedical use of OxyContin® across all of the possible combinations of answers to question PR04A. Therefore, remapping of the responses to these "enter all that apply" questions as part of the processing of transmitted data involved reassigning answers so that a non-missing value in a given variable had one, and only one, meaning across all respondents, regardless of the number of answers that respondents chose from a list or the order in which respondents keyed their answers. For example, a discrete variable was created as part of the remapping process that captured all reports of nonmedical use of OxyContin® from respondents who chose this drug as part of any of their answers to question PR04A. If respondents did not report nonmedical use of OxyContin®, the remapped variable for OxyContin® was assigned a code of 98 (i.e., blank).

In addition to choosing one or more applicable responses from a list, respondents could use function keys to answer "don't know" or "refused" as their first response to these "enter all that apply" types of questions. In situations where respondents answered "don't know," it would be reasonable to infer that the respondent did not know which particular item on the list applied to him or her. For example, if a respondent answered question PR04A as "don't know," this was inferred to mean that the respondent did not know whether he or she had ever misused codeine, Demerol®, Dilaudid®, and so on, through Ultram®. In this case, a code of "don't know" was propagated to each of the recoded "enter all that apply" variables as part of the daily processing

of the transmitted data. Similarly, if a respondent *refused* to answer question PR04A, a refusal code was propagated to all of the recoded variables on that list as part of the daily processing of the transmitted data.

2.2.3 Identification of Usable Cases

Once standard missing data codes had been assigned and the responses to "enter all that apply" questions had been remapped, the third key step in the preliminary processing of transmitted NSDUH data established the minimum item response requirements necessary for cases to be retained for weighting and further analysis (subsequently referred to as "usable" cases). These rules were designed to eliminate cases with unacceptable levels of item nonresponse (i.e., missing data), thereby retaining cases with lower levels of missing data and reducing the amount of statistical imputation needed for any given record. In addition, requiring lifetime use or nonuse to be fully defined for at least one substance (i.e., no missing data) can allow data for that substance to be used in statistical imputations for other drugs with missing data for lifetime use or nonuse.

In order for a case to be considered usable in NSDUH, both of the following requirements must be met.[5] The term "gate question" is used in connection with the usable case criteria because an affirmative response to these questions (e.g., "Have you ever, even once, used marijuana or hashish?") opens the "gate" to a series of other questions on use of the drug, and a negative response closes the "gate" and leads to the skipping of all other questions on use of that drug.

1. The lifetime cigarette gate question CG01 had to have been answered as "yes" or "no" so that lifetime use or nonuse could be fully defined for at least one substance. Data about lifetime use or nonuse of cigarettes is used in subsequent statistical imputations for other drugs where lifetime use/nonuse of other drugs is undefined.
2. At least 9 of the following 13 additional gate questions had to have answers of "yes" or "no": (1) chewing tobacco, (2) snuff, (3) cigars, (4) alcohol, (5) marijuana, (6) cocaine (in any form), (7) heroin, (8) hallucinogens, (9) inhalants, (10) pain relievers, (11) tranquilizers, (12) stimulants, and (13) sedatives.[6]

For cigarettes through heroin, respondents are asked a single "yes/no" question for their lifetime use or nonuse. Respondents who initially refuse to answer the gate question are asked a follow-up question to encourage them to reconsider their refusal. The usability criterion for these substances is met if these respondents change their initial refusal to an answer of "yes" or "no" (i.e., they provide a response that would no longer be considered to be a missing value).

For hallucinogens through sedatives, lifetime use or nonuse for the overall category is determined by asking multiple "yes/no" questions about lifetime use or nonuse of specific drugs

[5] The historical background and considerations for establishing usable case rules for the CAI data are discussed in a methodological chapter on editing the 1999 CAI data (Kroutil & Myers, 2002).

[6] Crack cocaine was not included in the usable case rule because the logic for asking about crack cocaine was dependent upon the respondent having answered the lifetime cocaine use question as "yes." In addition, although the CAI instrument asked about pipe tobacco, this was not included in the usable case rule because there was only one other question about pipe tobacco in addition to the gate question.

within the broader category (e.g., LSD within hallucinogens). Consequently, these questions are referred to as "multiple" gate questions. If any of these multiple gate questions are answered as "yes," then the respondent logically is a lifetime user for the overall category (e.g., hallucinogens).

For these multiple gate drug categories, the criterion for usability was considered to have been met if at least one lifetime gate question in the series was answered as "yes" or "no" (e.g., for hallucinogens, if at least one question in the series LS01A through LS01H was answered as "yes" or "no"). This rule was adopted for the multiple gate drug categories because requiring lifetime use or nonuse to be known for the overall category would have placed a more stringent requirement for usability on data from nonusers than from users. Specifically, unambiguous identification of lifetime nonusers for the overall category required them to answer "no" to *every* gate question in the multiple gate series because respondents could have been lifetime users of drugs that had missing data. Consequently, respondents who answered some multiple gate questions as "no" and also had some responses of "don't know" or "refused" would fail a usability rule for multiple gate drugs that required lifetime use or nonuse for the overall category to be known unambiguously. In contrast, respondents' status as lifetime users for the overall category was known if they answered "yes" to at least one drug in the series, even if they had given responses of "don't know" or "refused" for other questions in the series.

The types of follow-up questions that were administered in response to initial refusals varied for these modules with multiple gate questions.

- For hallucinogens, respondents were administered individual follow-up questions after refusals to report lifetime use or nonuse of the specific hallucinogens LSD, PCP, and Ecstasy. However, respondents were not asked a follow-up question to determine lifetime use or nonuse of any hallucinogens (i.e., regardless of which specific ones) if they refused to answer all gate questions for hallucinogens (including continued refusal to report lifetime use or nonuse of LSD, PCP, or Ecstasy).
- For inhalants, pain relievers, tranquilizers, and sedatives, respondents who refused to answer all gate questions in a module were asked a follow-up question to determine lifetime use[7] or nonuse of any drugs in the overall category. Respondents were not asked follow-up questions if they refused to answer a specific gate question (e.g., lifetime nonmedical use of Vicodin®, Lortab®, or Lorcet® in question PR03 for pain relievers) but they did not refuse to answer all gate questions in that module.
- For stimulants, respondents were administered a follow-up question if they refused to answer the gate question about lifetime nonmedical use of methamphetamine, Desoxyn®, or Methedrine®. Respondents also were administered a follow-up question to determine lifetime nonmedical use or nonuse of any stimulants if they refused to answer all gate questions for stimulants (including a refusal to answer the question about methamphetamine on follow up).

Despite these variations in how refusal follow-up questions were administered in these multiple gate drug modules, the usability principle that was described previously for modules

[7] For pain relievers, tranquilizers, and sedatives, this refers to nonmedical use.

with multiple gate questions applied to these follow-up questions: if respondents refused to answer a question (or series of questions), but then answered "yes" or "no" to the follow-up probe, then they were considered to have met the usability criterion for that module. In particular, if a respondent changed a refusal for lifetime use of Ecstasy to an answer of "yes" or "no," then the respondent was considered to have met the usability criterion for hallucinogens, regardless of whether he or she had missing data for other gate questions in the hallucinogens module.

Table 2.1 lists the follow-up questions in the core modules that were administered when respondents initially refused a gate question, including follow-up questions that were administered when there was a single gate question (i.e., for cigarettes through heroin) and follow-up questions that were administered in modules with multiple gate questions. Table 2.1 also lists the implications for the usable case criteria according to how respondents answered these follow-up questions.

Table 2.1 Effects on the Potential Usable Status of a Case Based on Responses to Follow-Up Questions for Refusals to Gate Questions in the Core Drug Modules

Module or Drug (if Applicable)	Question Number for Follow-Up Question	Consequence if Follow-Up Response Is "Yes" or "No"	Consequence if Follow-Up Response Is DK or REF
Cigarettes	CGREF1	Meets usable case criterion for cigarettes	Not a usable case
Snuff	CGREF3	Meets usable case criterion for snuff	Does not meet usable case criterion for snuff[1]
Chewing Tobacco	CGREF2	Meets usable case criterion for chewing tobacco	Does not meet usable case criterion for chewing tobacco[1]
Cigars	CGREF4	Meets usable case criterion for cigars	Does not meet usable case criterion for cigars[1]
Alcohol	ALREF	Meets usable case criterion for alcohol	Does not meet usable case criterion for alcohol[1]
Marijuana	MJREF	Meets usable case criterion for marijuana	Does not meet usable case criterion for marijuana[1]
Cocaine	CCREF	Meets usable case criterion for cocaine	Does not meet usable case criterion for cocaine[1,2]
Heroin	HEREF	Meets usable case criterion for heroin	Does not meet usable case criterion for heroin[1]
Hallucinogens/LSD	LSREF1	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"
Hallucinogens/PCP	LSREF2	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"

Table 2.1 Effects on the Potential Usable Status of a Case Based on Responses to Follow-Up Questions for Refusals to Gate Questions in the Core Drug Modules (continued)

Module or Drug (if Applicable)	Question Number for Follow-Up Question	Consequence if Follow-Up Response Is "Yes" or "No"	Consequence if Follow-Up Response Is DK or REF
Hallucinogens/Ecstasy	LSREF3	Meets usable case criterion for hallucinogens	Does not affect usable case status for hallucinogens if another gate question (or follow-up question) is answered as "yes" or "no"
Inhalants	INREF[3]	Meets usable case criterion for inhalants	Does not meet usable case criterion for inhalants[1]
Pain Relievers	PRREF[3]	Meets usable case criterion for pain relievers	Does not meet usable case criterion for pain relievers[1]
Tranquilizers	TRREF[3]	Meets usable case criterion for tranquilizers	Does not meet usable case criterion for tranquilizers[1]
Stimulants/Methamphetamine	STREF1	Meets usable case criterion for stimulants	Does not affect usable case status for stimulants if another gate question (or follow-up question) is answered as "yes" or "no"
Stimulants	STREF2[3,4]	Meets usable case criterion for stimulants	Does not meet usable case criterion for stimulants[1]
Sedatives	SVREF[3]	Meets usable case criterion for sedatives	Does not meet usable case criterion for sedatives[1]

DK = don't know; REF = refused.

[1] Overall status as a usable case will still be met if the usable case criteria are met for a sufficient number of other modules.

[2] The interview includes a follow-up probe if the gate question for crack cocaine is refused (CKREF) but crack cocaine is not included in the criteria for identifying usable cases.

[3] Follow-up probe is asked if the respondent refused to answer all gate questions for that module.

[4] Respondents who are routed to STREF2 also had refused the methamphetamine follow-up question STREF1.

The usable case rule was a necessary, but not sufficient, requirement for a case to be considered a final respondent. Cases that had sufficient data to meet the usable case criteria could still be treated as nonrespondents if their interview data appeared to be of poor data quality, as evidenced by potential response pattern problems (see Section 2.3.2).

2.2.4 Creation of Cumulative Quarterly Unedited Data Files

Following daily processing of the data, each day's SAS dataset was merged with the cumulative data that had been transmitted up to that point in the quarter. At the end of the quarter, a complete data file was produced that contained all cases that had been transmitted during the quarter. Each quarterly data file then underwent additional initial cleaning and processing (prior to the editing procedures) to modify or correct field errors, such as erroneous ID entries by the field staff. The cleaned-up (but otherwise unedited) SAS datasets from the first two quarters (also known as 6-month data) and from all four quarters (also known as 12-month

data) served as the usual starting points for the subsequent logical editing procedures that are described in Chapters 4 through 11.[8]

2.3 Preliminary Editing and Coding of Processed Interview Data

In addition to procedures that were described in Section 2.2 following receipt of transmitted data from the field, preliminary coding and processing of unedited interview data encompassed the following activities:

- coding of "OTHER, Specify" data (Section 2.3.1);
- investigation of response patterns in records that otherwise met the usable case criteria (Section 2.3.2);
- edits to "date-dependent" variables (if applicable) when the interview date was judged to be questionable (Section 2.3.3).

The first two of these activities could occur or did occur prior to creation of the cumulative quarterly unedited data files (Section 2.2.4). Processing of age-related variables was performed once the cumulative unedited data file from the first two quarters had been created and then again when the cumulative unedited data file from all four quarters had been created. Edits to date-dependent variables were not performed until final interview dates had been created for respondents (see Section 4.2.1 in Chapter 4).

Note that a code to denote "bad data" (i.e., to indicate an inconsistency or some other problem in the original data) could be assigned to variables during any of these activities, except processing of age-related variables. The following codes were assigned to denote "bad data": 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned. Codes for bad data were treated as missing values. Any assignment of bad data codes was done in subsequent editing steps, not as part of the nightly processing of transmitted data.

2.3.1 Coding of "OTHER, Specify" Data

This activity took alphanumeric (text) answers that respondents or interviewers had typed (e.g., specific other drugs used, specific other payment sources of treatment) and converted them into numeric codes. These alpha answers (and the numeric codes resulting from them) are referred to as "OTHER, Specify" data.

Coding of the "OTHER, Specify" variables was accomplished through computer-assisted procedures.[9] "OTHER, Specify" responses were first converted to all capital letters because

[8] Edits are run on preliminary data from the first two quarters of a given survey year to identify any updates that need to be made to the programs for use with the full data from all four quarters. Running the edits on data from the first two quarters is particularly useful for testing the programs for any sections of the CAI instrument that are new or have changed since the previous survey year.

[9] A system has been in place since 2002 for the daily coding and processing of the "OTHER, Specify" variables for drugs and tobacco brands. This system now encompasses the daily coding and processing of *all* "OTHER, Specify" variables from the survey that underwent assignment of numeric codes, except for codes pertaining to the industry in which respondents were employed and their current or former occupations. Coding of industry and occupation data was handled by the U.S. Census Bureau.

respondents could use different combinations of uppercase or lowercase characters to provide an otherwise identical response. If an exact match was found between what the respondent keyed and an entry in the data dictionary (e.g., "ALCOHOL"), the computer-assisted procedures assigned the appropriate numeric code (e.g., 807 for alcohol). The system could also accommodate commonly encountered misspellings (e.g., "ALCHOHOL").

Typed answers that the respondent provided that did not match an existing data dictionary entry were reviewed by analysts on a flow basis during the quarter to determine whether an existing code should be assigned to the response or a new code should be created. Based on these decisions, new entries were added to the relevant dictionaries on a quarterly basis—including new dictionary entries corresponding to existing codes—for use in daily processing of data that were transmitted from the field in subsequent quarters. Analysts could also decide not to add a particular response to the data dictionary, in which case the response would be output for review and coding on a case-by-case basis if it was reported in a future quarter or survey year. In addition, analysts reviewed the codes that were assigned through the computer-assisted process to verify that codes were being assigned correctly. Over time, these procedures have reduced the turnaround time and burden on analysts for producing the coded "OTHER, Specify" data.

Particularly for drugs and tobacco brands (Sections 2.3.1.1 and 2.3.1.2, respectively), which used the same data dictionaries for a number of "OTHER, Specify" variables, matching a written response to a numeric code in the data dictionary would result in that code being assigned *no matter where* respondents typed that answer. For example, a given entry in the data dictionary for drugs would be assigned the same code if it appeared in the hallucinogens module one year and appeared in the pain relievers module another year. Similarly, an entry in the data dictionary for tobacco brands would be assigned the same code regardless of the type of tobacco where the response appeared. In some situations, however, the same response could have different meanings depending on the context. For example, a given tobacco brand name with no other associated information could apply to a brand of cigarettes or to a brand of cigars, depending on whether it was specified as a cigarette brand or as a cigar brand.

As with the "enter all that apply" data that were discussed previously, respondents could answer the "OTHER, Specify" questions as "don't know" or "refused," which were then reassigned to the respective codes of 9994 or 9997, as described in Section 2.2.1. Respondents could also type in an equivalent response to "don't know" (e.g., "no idea") or "refused" (e.g., "too personal"). These equivalent responses were assigned the relevant codes for missing data (Section 2.2.1) as part of the coding procedures. For typed responses that were nonsensical or otherwise nonresponsive to the request to specify additional information, codes for bad data (Section 2.3) were assigned.

The remainder of the discussion in this section focuses on issues related to coding of the "OTHER, Specify" data according to the type of other information that was requested from respondents:

- other drugs (Section 2.3.1.1),
- other tobacco brands that respondents used most commonly in the past 30 days (Section 2.3.1.2),

- other race or ethnicity (Section 2.3.1.3), and
- additional "OTHER, Specify" data in noncore sections (Section 2.3.1.4).[10]

Except for mentions of other drugs (see Section 2.3.1.1), "OTHER, Specify" data in the 2012 NSDUH typically were intended to capture a single "other" response (e.g., most important other reason for not receiving mental health treatment in the past 12 months). If respondents typed in responses for which multiple codes could apply (e.g., multiple reasons for not receiving mental health treatment instead of the single most important reason), the standard procedure was to assign a code to the first response that could be coded. Relevant sections elsewhere in this report indicate when exceptions were made to this more general approach (e.g., coding of "OTHER, Specify" data for youth mental health service utilization, described in Section 7.4.16).

2.3.1.1 Other Drugs

In the core modules for hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives, respondents could report lifetime use or nonmedical use of drugs other than those that were specifically asked about in the respective modules. Respondents also could report lifetime injection of drugs other than heroin, cocaine, methamphetamine, or other stimulants in the noncore special drugs module. In the noncore substance treatment module, respondents could report current or most recent treatment or the perceived need for treatment in the past 12 months for drugs other than those for which they reported lifetime use in the core modules.

In the "OTHER, Specify" drug questions for both core and noncore modules, respondents could enter up to five responses (in five data entry fields) for other drugs. For the most part, respondents specified a total of only one drug in these questions or they specified only one drug in each field. Neither of these situations required any special handling of the data.

However, additional procedures were required to handle the following situations in the "OTHER, Specify" drug data.

1. Respondents specified more than one drug within a single data entry field but they specified a total of five or fewer drugs across the five available fields.
2. Respondents specified a total of more than five substances across the five available data entry fields (i.e., by definition, they specified multiple drugs within one or more fields).

In the first situation, codes for the additional drugs were moved to the next unused "slot(s)" (i.e., variable[s]). For example, if a total of three drugs were specified, with two of them being specified in the first field and the third being specified in the second field, the additional code from the first field was moved to the third "OTHER, Specify" variable. Consequently, the responses in the individual "OTHER, Specify" drug variables did not always correspond to the order in which respondents reported use of these drugs.

[10] See Section 1.1 and Tables 1.1 and 1.2 in Chapter 1 for details about the core and noncore structure of the NSDUH questionnaire and contents of the core and noncore sections.

When respondents specified more than five substances in the available fields, duplicate mentions of drugs were identified and removed. Duplicate mentions could include the same drug being mentioned more than once in the "OTHER, Specify" data or a drug being reported in the gate questions for a given module and also being reported in the "OTHER, Specify" data (e.g., if question LS01a for LSD was answered as "yes" and LSD also was reported in the "OTHER, Specify" data for hallucinogens). If more than five unique mentions of drugs remained after any redundant mentions had been eliminated, further priority in retaining responses in the final drug codes was given to (1) mentions of illegal or prescription-type drugs, as opposed to "over-the-counter" (OTC) drugs that are legally available without a prescription; and (2) mentions that were relevant to the category of interest (e.g., in the hallucinogens section, giving priority to mentions of hallucinogens over mentions of drugs that were not classified as hallucinogens).

In rare instances, the procedures described previously still did not yield five or fewer unique drugs in the "OTHER, Specify" data for a given module. In these situations, the highest priority was given to retaining codes for drugs that had no approved medical use in the United States or were deemed to have a greater potential for dependence or abuse, and the lowest priority was given to retaining codes for drugs that were not classified as controlled substances.[11] In the extremely rare situations where five or more unique drugs remained after all of these measures had been applied, the first five remaining "OTHER, Specify" codes were retained and mentions of any additional drugs were dropped from the coded data that were available for further editing or analysis.

2.3.1.2 Other Tobacco Brands

The CAI instrument included questions to identify the specific brands of tobacco that were used most commonly by respondents who reported use in the past 30 days of cigarettes, chewing tobacco, snuff, or cigars. For these four types of tobacco products, respondents who reported use in the past 30 days could choose from a list of brands shown on the computer screen[12] or they could indicate use of "a brand not on this list." Respondents who gave the latter answer were asked to type in the name of the specific other brand that they used.

The basic coding scheme for the "OTHER, Specify" tobacco brand categories was as follows.

- Codes of 101-199 and 1001-1999 were reserved for cigarette brands.
- Codes of 201-299 and 2001-2999 were reserved for chewing tobacco brands.
- Codes of 301-399 and 3001-3999 were reserved for snuff brands.
- Codes of 401-499 and 4001-4999 were reserved for cigar brands.
- Codes of 501-599 and 5001-5999 were reserved for pipe tobacco brands.
- Codes in the 600-699 and 700-799 series were reserved for miscellaneous tobacco and nontobacco responses.

[11] The drug scheduling classifications, Schedules I through V, under the Controlled Substances Act were used in making these determinations. See www.deadiversion.usdoj.gov/21cfr/21usc/index.html and www.deadiversion.usdoj.gov/schedules/index.html for details.

[12] For cigarettes, the listing of brands was split between two different computer screens.

Codes were assigned to the "OTHER, Specify" tobacco brand data according to these categories, regardless of whether the response came from the section for cigarettes, snuff, chewing tobacco, or cigars within the tobacco module. This coding scheme was particularly relevant for the smokeless tobacco data for snuff and chewing tobacco, where snuff brands could be reported as the brand of chewing tobacco that was used most often in the past 30 days, or vice versa. Thus, if a respondent specified a brand of snuff in the chewing tobacco section, the "OTHER, Specify" response for the chewing tobacco brand was assigned a code in the 300 or 3000 series for snuff brands. Similarly, if a respondent specified that the brand of other cigarettes that he or she smoked most often in the past 30 days was actually a brand of little cigars, the "OTHER, Specify" response for cigarette brands was assigned a code in the 400 or 4000 series for cigars.

Note that the coding for a particular tobacco brand did not capture further details, because the main aim in the coding was to capture information about any use within a particular brand label. For example, the code for a particular brand did not capture details such as length (e.g., for cigarettes), size or shape (e.g., for cigars), or flavor varieties.

2.3.1.3 Other Race or Ethnicity

In the interviewer-administered core demographics module (see Chapter 4), NSDUH respondents were asked about their Hispanic/Latino origin and race information in QD04 and QD05, respectively. If respondents reported in QD05 that they were Asian, they were asked in QD05ASIA to report which Asian group best described them. However, they could consider that the categories presented to them for Hispanic/Latino origin, race, or Asian ethnicity did not apply to them. In these situations, respondents reported their "other" Hispanicity, race, or Asian ethnicity to the interviewers, who then typed in the respondents' answers.

The computer-assisted coding procedures and use of data dictionaries that were described in Section 2.3.1 applied to these "OTHER, Specify" data for race and ethnicity as well. In most cases, new responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican."

Regardless of the source (QD04, QD05, QD05ASIA) for these "OTHER, Specify" data, the write-in responses were used in subsequent editing of Hispanicity and race to determine the respondents' final Hispanicity and race (see Sections 4.2.7, 4.3.2, 4.3.3, and 4.3.4). Thus, in coding the "OTHER, Specify" data, each write-in was assigned two codes, one for race and the other for Hispanicity. If an interviewer entered both a geographical entity and a group within a particular race in the "OTHER, Specify" response, such as "Black Cape Verdean," the geographical entity was ignored in the race code and the respondent was coded as "Black/African American" for the race code. The geographic information Cape Verde was captured in the Hispanic/Latino code for "Cape Verde."

2.3.1.4 Additional Noncore "OTHER, Specify" Data

There were three types of "OTHER, Specify" questions in the noncore self-administered or interviewer-administered sections:

- those where the "OTHER, Specify" item was a follow-up to a lead question that typically was answered as "yes" or "no." Depending on the nature of the lead question, either an affirmative or a negative response to the lead question could govern whether respondents were asked to specify something;
- those where the "OTHER, Specify" item was a follow-up to a response category for "other" in an "enter all that apply" question; and
- those where respondents did not get the opportunity to choose the "other" response (and specify something) if they already chose another category from the list.

Coding of these noncore "OTHER, Specify" variables was performed according to the general principles described in Section 2.3.1. Otherwise, minimal additional decision making was involved in assigning codes to the responses for these variables. Descriptions of these variables are included here for completeness.

An example of the first type of "OTHER, Specify" question is question TX42JSP in the noncore self-administered substance treatment module (i.e., specify other source that paid/will pay for the last or current substance treatment). Only those respondents who reported in question TX42J that some other source paid for their last substance treatment or counseling or that some other source would pay for their current treatment (TX42J = 1) were routed to TX42JSP and asked to report the other payment source. Respondents could report other sources of payment for their treatment in questions TX42A to TX42I and also report some other payment source in TX42J.

Question ADMT15 in the noncore self-administered adult mental health service utilization module (i.e., specify the other location where outpatient mental health treatment was received in the past 12 months) is an example of the second type of "OTHER, Specify" question. Adult respondents aged 18 or older could report in question ADMT14 that they received mental health treatment or counseling in the past 12 months in up to six different types of outpatient settings, including "some other place." Adults were routed to ADMT15 if they reported receiving outpatient treatment in some other place in ADMT14. Because ADMT14 was an "enter all that apply" question, respondents could choose any of the specific locations that were listed in the question (e.g., an outpatient mental health clinic or center) and also report receipt of mental health treatment in some other place.

The third type of "OTHER, Specify" question is represented by the following three questions:

- SD16SP, which was associated with question SD16 (how respondents obtained their last needle for injecting drugs) in the self-administered special drugs module;
- TX25SP, which was associated with question TX25 (main place where the respondent received or was currently receiving substance abuse treatment in the past 12 months) in the self-administered substance treatment module;[13] and

[13] Although question TX04ISP (specify other location where the respondent received treatment for alcohol or other drug use in the past 12 months) was preceded by a "yes/no" question, TX04ISP used the same codes as TX25SP to allow the same data dictionary to be used for processing both of these "OTHER, Specify" variables (see Section 2.3.1).

- QD24SP, which was associated with question QD24 (reason for leaving school without getting a high school diploma) in the interviewer-administered noncore education module.

Respondents first were presented with a list of options in the "lead" question (i.e., SD16, TX25, or QD24), including an option for "other" (e.g., other reason in QD24). If respondents chose any response from the list of options in the lead question except for "other," they were *not* routed to the "OTHER, Specify" question. Rather, data from the lead question and the specify question were combined into a single final variable.[14] When respondents chose the other category in the lead question, but they specified something that was coded as a missing value (i.e., don't know, refused, bad data, or blank), then the final edited variable (e.g., GNNDGET) retained a code corresponding to other, as opposed to being assigned a missing value.

The edits applied to GNNDGET, TXLTYMN, and LFSCHWHY (including similar types of questions that could be added in the future) were designed to provide analysts with a *standardized* way to readily identify when it could be logically inferred that respondents should have chosen a given response option from the preceding question (i.e., rather than "other"). For GNNDGET as an example, codes of 1 to 4 applied to answers that respondents gave directly from question SD16 (e.g., 1 = Bought the needle from a pharmacy). Category 5 in SD16 was "Got the needle some other way." Although the coding sequence for "OTHER, Specify" responses in question SD16SP could have resumed at number 6 for responses corresponding to "Bought the needle from a pharmacy," assigning a code of 6 for "OTHER, Specify" responses that corresponded to category 1 in question SD16 would not enable analysts to readily see the logical connection between the "OTHER, Specify" response and the available response choice in the question that preceded it.

In this example, an alternative to enable analysts to see the logical connection between reports of buying the needle from a pharmacy in SD16 and corresponding reports in SD16SP would be to assign a code of 11 to responses in SD16SP that corresponded to category 1 in SD16. However, this coding scheme cannot be applied to question QD24, because QD24 listed 15 possible reasons for leaving school, not including "other reason." A code of 11 could not be used in identifying responses in QD24SP that corresponded to category 1 in QD24 ("School was boring or I didn't want to be there"), because that code was reserved in the edited variable LFSCHWHY for responses of 11 in QD24 (i.e., Moved here from another country and didn't enroll [or dropped out] because of language or other problems).

In consideration of this issue, responses in these "OTHER, Specify" variables that corresponded to existing response categories in their respective lead questions were coded starting with the number 21, with the coding proceeding in the order of the existing response categories. This procedure prevented overlap of "OTHER, Specify" codes with available responses when lead questions had 10 or more available choices, as was the case with QD24, and offered analysts a standardized way to interpret these values. For example, if analysts wanted to treat codes of 21 to 24 in GNNDGET that came from "OTHER, Specify" data as being equivalent in an analysis to the corresponding codes of 1 to 4, the analyst could simply subtract

[14] The final edited variables in 2012 were GNNDGET (based on data from SD16 and SD16SP), TXLTYMN (based on data from TX25 and TX25SP), and LFSCHWHY (based on data from QD24 and QD24SP).

20 from any codes with values of 21 to 24 to recode these values to the corresponding codes of 1 to 4; the same procedure could be applied if analysts wanted to treat codes of 21 to 35 in LFSCHWHY as being equivalent to codes of 1 to 15.

For similar reasons, the coding sequence for responses in these "OTHER, Specify" variables that did not correspond to responses from the list in the lead question resumed at number 41 (e.g., for GNNDGET, 41 = Given by/stolen from friend/acquaintance of friend/nonrelative). Although codes for responses that did not correspond to available choices from a lead question could have resumed at a lower number—especially for SD16SP and TX25SP—this approach standardized the assignment of codes across these types of "OTHER, Specify" variables (i.e., new codes always started at 41), minimized the risk of overlap between codes for these two types of responses, and allowed flexibility if new response choices were added to a lead question in a future year.

2.3.2 Investigation of Response Patterns in Usable Records

Although conversion to CAI reduced or eliminated some data quality problems that could occur in a PAPI format, it also was recognized that the audio computer-assisted self-interviewing (ACASI) environment could encourage some respondents to use the computer keyboard to enter nonsensical patterns of answers if they were not paying attention to questions or were not taking the interview seriously for other reasons. Thus, even if a respondent had sufficient data to meet the usable case criteria described in Section 2.2.3, certain patterns of answers could call into question the overall validity of the respondent's data.

In response, a data diagnostics program was developed to screen for the following patterns of responses that might raise questions about the validity of the interview as a whole:

- high numbers of "yes" responses to lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives), which might indicate that respondents were indiscriminately keying data without paying attention to what they were entering;
- alternating "yes/no" responses to questions about lifetime use of specific hallucinogens, inhalants, or psychotherapeutics (or alternating patterns of "response entered/not entered" in the psychotherapeutics sections), which might indicate some type of pattern-making;
- high numbers of illegal drugs that respondents reported using every day or just about every day in the past year or past month (where applicable), in which case one might question either the validity of the answers or the respondent's competence to complete the interview;
- high numbers of substances that respondents reported first using at age 1 or 2, which might indicate indiscriminate keying of 1s or 2s, especially given that the age-at-first-use questions followed gate questions where a response of 1 denoted "yes" and a response of 2 denoted "no"; and

- consistent keying of the same code (either 1 or 2) throughout one or more modules, which would suggest a pattern of indiscriminate answering.[15]

These patterns of responses were examined on a case-by-case basis to determine whether a case should be retained as a final respondent or dropped.

Fewer than 10 cases in 2012 met the usable case criteria but were treated as nonrespondents because their responses were of questionable validity, based on one or more of the patterns described above. In addition, fewer than 10 cases in 2012 were retained as respondents, whose original responses to questions in one or more core drug modules were replaced with bad data codes. This process included setting responses to bad data that indicated they were lifetime users of a given drug. For example, data for some respondents were set to bad data because the respondents keyed values of 1 or 2 to every question that was asked in a module, beginning with the age-at-first-use question. Data for the lifetime variables for these cases were set to bad data as part of the edits for the lifetime drug use variables (see Section 6.2.1). For example, if a case was identified that had "bad" stimulants data, the lifetime stimulants data corresponding to responses in questions ST01 through ST05 were set to bad data as part of the lifetime drug use edits, and a flag was set to indicate that data subsequently needed to be set to bad data for related variables pertaining to nonmedical use of prescription-type stimulants (both core and noncore variables pertaining to stimulants).

2.3.3 Edits to "Date-Dependent" Variables

The CAI instrument used the interview date information that was stored by the computer to create "date fills" during the interview that indicated the starting dates for questions pertaining to the "past 30 days" and "past 12 months." Specifically, the starting date for the past 30-day (or 12-month) period was calculated as exactly 30 days (or exactly 12 months) prior to the stored interview date.[16] Thus, in the recency-of-use questions that asked respondents when they last used the drug of interest, the response category "within the past 30 days" included a date fill to remind respondents when the past 30-day reference period began for them. Similarly, introductions to specific questions about frequency of use of a particular drug in the past 12 months and past 30 days included date fills to remind respondents of the period they should be thinking about when answering these questions.

---

[15] An important change to the CAI instrument since 2001 is that response categories for certain consistency checks use 4=yes/6=no instead of the format of 1=yes/2=no that was used prior to 2001. For example, if a respondent reported first using marijuana at age 1 or 2, the respondent could not use a response of "1" or "2" in question MJCC05 to verify that this age at first use was correct. This change to the CAI instrument was designed to stop respondents if they had been engaged in a pattern of keying responses of "1" or "2."

[16] If a respondent broke off and then resumed the interview at a later date, the "fills" in the interview that defined the past 30-day and past 12-month reference periods were updated for remaining questions when the interview resumed. This was done because 30 days and 12 months from the date when an interview resumed might be later than the 30-day and 12-month periods when a respondent had previously been asked questions. Thus, the "fills" that defined these reference periods during the interview were dynamic rather than static. If a respondent resumed the interview and went back to a section that he or she had completed prior to the breakoff, the date "fill" would be reset based on the new date when the interview resumed. In practice, however, the tendency is to proceed with the parts of the interview that have not been completed rather than to go back to earlier sections.

Data that could be affected by questionable interview dates were edited as needed once the edited interview date variable INTDATE had been created for all final respondents (Section 4.2.1). As part of the procedures for creating INTDATE, an indicator variable (EIIDATE) was created that specified how the final interview date was assigned. For example, EIIDATE could indicate that the interview date that had been stored by the CAI program during the interview had an incorrect year or was outside of the quarter in which the interview had been completed.

With the exception of 2008, however, no cases since 2004 have had an originally entered interview date that was sufficiently problematic to call into question the respondent's answers to date-dependent questions in self-administered sections of the interview. If this problem were to occur, subsequent data in the self-administered modules that were dependent on the interview date would be considered problematic. For example, if the CAI program calculated a 30-day reference period based on an incorrect interview date, answers that the respondent gave on the number of days that he or she used different drugs in the past 30 days could reflect use of these drugs in a period other than the intended 30-day reference period.

This interview date issue did not present a problem if respondents never used a particular drug (or had never engaged in other behaviors). Moreover, if the respondent reported never having engaged in a particular behavior, the CAI program skipped that person out of questions where the interview date would be important for establishing reference periods. For these reasons, some cases where there was some question about the interview date could still be retained as final respondents.

Instead of cases being dropped, the edits that are in place would set the following types of self-administered questions to bad data if respondents were routed to them:

- questions pertaining to behaviors in the past 30 days;
- questions pertaining to behaviors in the past 12 months;
- questions pertaining to the most recent time that an event occurred (e.g., when a respondent last used a drug of interest); and
- questions pertaining to the respondent's age when an event occurred (e.g., the age when the respondent first used a drug of interest).

Self-administered questions about age at first use and ages when other events occurred were not related directly to the interview date but were related indirectly via respondents' ages; that is, respondents' ages were calculated by comparing the date of birth with the interview date. In turn, age-at-first-use and other age data in the self-administered modules were compared for consistency with the respondents' ages.

For the types of questions that were described previously, respondents' answers would be set to bad data before any further editing was done. Setting the responses to bad data prior to any subsequent editing allowed analysts during logical editing to distinguish between situations where the data were deemed to be bad prior to any further editing being done and situations where a variable might be set to bad data in subsequent editing steps because of inconsistencies with other data in the respondent's record (see Section 2.4.3).

This edit would not apply to gate questions that asked whether a respondent had ever engaged in a behavior (e.g., "Have you ever smoked part or all of a cigarette?"). As noted above, whether the respondent had ever engaged in a particular behavior prior to being interviewed is not dependent on the value stored for the interview date. This edit also does not apply to questions within a module that asked whether a more detailed behavior of interest was *ever* true for this respondent (e.g., "Have you smoked at least 100 cigarettes in your entire life?"). The decision also was made that this edit would not apply to sections of the interview that were administered directly by the field interviewers (FIs).

2.4 General Issues and Principles for Editing NSDUH Data

Previous sections discussed broader data processing and coding procedures in the data as a whole or in large sections, such as the core section of the interview. In contrast, this section focuses on the general principles of data processing and editing across and within modules, once the initial processing of transmitted interview data had been completed (i.e., the procedures described in Section 2.2) and after cases with patterned responses or questionable interview date information had been identified (Section 2.3).

The following specific issues are addressed in this section:

- editing across modules (Section 2.4.1);
- assignment of relevant "not applicable" codes (Section 2.4.2);
- additional assignment of NSDUH missing data codes (Section 2.4.3);
- assignment of codes to edited "enter all that apply" variables (Section 2.4.4);
- "back-editing" based on "OTHER, Specify" data (Section 2.4.5); and
- general principles for the "flag and impute" procedures for core drug variables (Section 2.4.6).

2.4.1 Editing across Modules

An important principle that was followed in editing the data was that responses from one module (e.g., hallucinogens) generally were not used to edit variables in another module (e.g., inhalants).[17] In particular, the noncore self-administered data (special drugs module through the consumption of alcohol module) were generally *not* used to edit related variables in the core self-administered modules (tobacco through sedatives).[18] For example, if a respondent reported in the core heroin module that he or she last used heroin more than 12 months ago, but then reported in the noncore special drugs module that he or she last used heroin with a needle more than 30 days ago but within the past 12 months, then the core heroin recency variable HERREC was *not* edited to take into account this more recent indication of heroin use from the noncore module. Rather, HERREC retained the respondent's initial response that he or she last used heroin more than 12 months ago. Consequently, the documentation for the noncore variables in NSDUH

[17] See Section 1.1 in Chapter 1 for a definition of modules.

[18] An exception to this principle that is discussed in Chapter 6 involved the editing of core data on use of methamphetamine to incorporate questions about methamphetamine use that have been included in the noncore special drugs module since 2006.

codebooks includes a footnote to indicate that these variables may be inconsistent with data in other sections of the interview.

This principle of not using noncore data to edit core data was important for maintaining consistent data to assess trends in substance use. If variables in core modules were allowed to be edited based on respondents' answers in the noncore modules, key drug use estimates could change across years as noncore questions or modules were added or deleted.

One exception to this principle involved situations in which responses in one or more modules governed whether respondents were asked questions in another module. For example, the substance treatment module was relevant only for respondents who reported some lifetime use of alcohol or other drugs, excluding cigarettes. Respondents who reported in the core modules that they had never used alcohol, illicit drugs, or prescription-type psychotherapeutics for nonmedical reasons (i.e., pain relievers, tranquilizers, stimulants, or sedatives) were not asked corresponding follow-up questions in the substance treatment module. In such cases, during the editing process, blank values in the substance treatment variables were replaced with codes to indicate that respondents were not asked the follow-up questions in this module because they reported never having used any of the relevant core drugs.

2.4.2 Assignment of Relevant "Not Applicable" Codes

Because the CAI logic controlled whether respondents were skipped out of some questions based on their answers to previous questions, an important aspect of editing the NSDUH data involved replacing missing data codes in the unedited data with appropriate codes to indicate that the questions had been skipped because they did not apply. The following codes were assigned when respondents were skipped out of a given question *and it could be determined unambiguously* that the respondent had legitimately been skipped out of the question, based on the answer(s) to one or more previous questions:

- 91 (or 991, or 9991, etc.) = NEVER USED [DRUG(s) OF INTEREST];
- 93 (or 993, or 9993, etc.) = USED [DRUG] BUT NOT IN THE PERIOD OF INTEREST; and
- 99 (or 999, or 9999, etc.) = LEGITIMATE SKIP.

Strictly speaking, codes of 91 and 93 in the CAI data could be considered variants of the more generic legitimate skip code. Their use was designed to provide analysts with more information about the reason that respondents were skipped out of a particular question.

Codes of 91 and 93 were used most often in the core drug sections of the interview. For example, codes of 91 (or 991, etc.) in the marijuana section denote the pattern where respondents were skipped out of all remaining marijuana questions because they answered "no" to the lifetime marijuana question MJ01. Similarly, codes of 93 in the marijuana section denote situations where respondents were lifetime marijuana users but were definitely not users in the past 30 days or past 12 months or both.

Codes of 91 and 93 also were used to a limited extent in noncore sections of the interview because the CAI logic took into account respondents' prior answers to core drug use questions to

determine whether particular noncore questions applied. For example, questions about cocaine in the substance dependence and abuse module were relevant to respondents who used cocaine in the 12 months prior to the interview. Thus, if a respondent last used cocaine more than 12 months prior to the interview, codes of 93 in the substance dependence and abuse variables pertaining to cocaine would signify to an analyst why the CAI program skipped the respondent out of these questions. Similarly, codes of 91 in the substance dependence and abuse variables for cocaine signified to an analyst that the respondent was skipped out of these questions because he or she had never used cocaine.

Legitimate skip codes of 99 were used most often in the noncore self-administered sections of the interview or in interviewer-administered sections. For example, the youth experiences module was intended to be administered only to respondents aged 12 to 17. If a respondent was 18 or older, codes of 99 were assigned in the editing process to the skipped youth experiences variables. Similarly, if a respondent had used alcohol or some other drug at least once in his or her lifetime, but answered the lifetime substance treatment question TX01 as "no," then the CAI program skipped the respondent out of all remaining substance treatment questions. Codes of 99 were assigned to the skipped substance treatment variables in this situation to signify that the respondent had used alcohol or drugs at least once but had never received substance abuse treatment.

The following codes also were assigned through editing:

- 81 (or 981, or 9981, etc.) = NEVER USED [DRUG(s)] Logically assigned;
- 83 (or 983, or 9983, etc.) = USED [DRUG(s)] BUT NOT IN THE PERIOD OF INTEREST Logically assigned; and
- 89 (or 989, or 9989, etc.) = LEGITIMATE SKIP Logically assigned.

These codes were given values in the 80s to signify that existing values were to be overwritten during the editing process. For example, if a respondent was somehow routed into the youth experiences module, but that respondent was subsequently classified as being 18 or older (see Section 4.2.2 in Chapter 4), then any answers that the respondent gave in the youth experiences module were overwritten with codes of 89 (or 989, etc.). These codes signify that this adult respondent logically was not eligible to be asked the youth experiences questions.

However, these codes were assigned *only* in situations where there was total certainty that a respondent should have skipped a question. For example, if a respondent reported in question MJ01 that he or she had never used marijuana, it was absolutely clear that subsequent questions about marijuana use (e.g., age at first use of marijuana) did not apply. See Section 3.1 for a description of imputation indicators, including a brief discussion of the codes assigned to respondents whose imputed values meant that they legitimately skipped out of the corresponding questions.

The CAI skip logic often treated responses of "don't know" or "refused" to gate questions as equivalent to a negative response. For example, if a respondent was uncertain whether he or she had ever used marijuana (and answered question MJ01 as "don't know"), the CAI program skipped the respondent out of all remaining questions about marijuana use, as though the

respondent had never used it. From the standpoint of respondent burden, this makes sense. There is little value in asking a respondent who did not know whether he or she had ever used marijuana, "How old were you the first time you used marijuana or hashish?" Implicit in this question is that respondents have used marijuana at least once in their lives.

Although the CAI program skipped respondents out of questions if they answered a gate question as "don't know" or "refused" (or gave similar answers on follow-up if they initially refused to answer a gate question), these types of responses to a gate question are ambiguous and do not provide an analyst with conclusive information one way or the other about the behavior or event of interest. Consequently, such responses could be thought of as *potentially* affirmative responses, as opposed to inferring that they are negative responses. In particular, as noted previously, respondents who initially refused to answer a question about their lifetime use or nonuse of a drug were given a second opportunity to answer the question as "yes" or "no." Similarly, if a respondent who initially did not know whether he or she had ever used a drug had thought about the issue further, the respondent may have recalled a time when he or she in fact had used it—and more detailed questions about use of the drug would have been relevant for this respondent. Alternatively, if the respondent gave more thought to the issue and decided that he or she really should answer the lifetime drug use question as "no," an analyst would have a solid basis for determining that subsequent questions did not apply.

Further, the procedures for statistically imputing missing data did not automatically infer lifetime nonuse when respondents provided ambiguous information about whether they had ever used a given drug. Rather, such respondents were eligible to be statistically imputed to be lifetime users or nonusers. For these reasons, variables retained missing values in the editing procedures when questions were skipped due to respondents answering a lead question as "don't know" or "refused" (or answering in a similar manner in response to a follow-up probe).

2.4.3 Additional Assignment of NSDUH Missing Data Codes as Part of Editing within a Module

Previous sections described the procedures for assigning missing data codes in the interview data as a whole or in large sections, such as the self-administered core and noncore sections, when the interview date was questionable (Section 2.3.3). This section discusses additional principles for assigning (or retaining) missing data codes as part of the editing procedures for a given module.

In particular, if respondents refused a single or multiple gate question or questions that governed a skip pattern in a module, refusal codes were assigned to all of the subsequently skipped items in the module as part of the editing procedures (i.e., the lead refusal was propagated); that is, it was logically inferred that the respondents were globally refusing to answer *any* questions on that topic.

This propagation did not occur when respondents answered a gate question or questions as "don't know." Rather, values of "blank" (no answer) were retained in the skipped questions. Unlike the situation for responses of "refused," it does not follow logically that a response of "don't know" to a gate question would imply that the respondent would answer "don't know" to all subsequent questions on that topic. For example, if a respondent answered the lifetime

marijuana question as "don't know," assigning a don't know code to the age-at-first-use variable (corresponding to question MJ02) would imply that the respondent was a lifetime user but did not recall the age when he or she first used.

In addition, data sometimes were identified that were inconsistent with other data in a respondent's record. For example, if a respondent reported first using a drug at an age greater than his or her current age, the CAI program indicated to the respondent that this age at first use was inconsistent. The respondent was prompted to revise the age at first use, his or her current age, or both, to make the data consistent. As noted in Chapter 1, however, respondents did not always resolve these types of inconsistencies. If the age at first use was still inconsistent with the respondent's age, despite the opportunity that the respondent had to resolve the inconsistency, then a bad data code was assigned to the age-at-first-use variable to indicate that the data were inconsistent with other data. As was the case with prior survey rounds, the following codes were assigned to denote "bad data" (i.e., usually inconsistent with other data): 85 (or 985, or 9985, etc.) = BAD DATA Logically assigned.

Other situations where bad data codes were assigned are discussed below in connection with specific steps in the editing process.

2.4.4 Assignment of Codes to Edited "Enter All That Apply" Variables

As noted in Section 2.2.2, the initial creation of separate variables for "enter all that apply" questions involved assignment of the relevant code that was shown to respondents during the interview for that question, or a code of 98 (blank) if respondents did not choose that item from the list; these procedures were implemented for "enter all that apply" questions anywhere in the entire interview. The additional edits described in this section were implemented within modules that had "enter all that apply" variables. In 2012, these variables were coded as 1 if respondents chose that item from the list, and values of 98 were recoded with a value of 6. Documentation for these edited values for "enter all that apply" variables was as follows:

- 1 = Response entered, and
- 6 = Response not entered.

For example, if a respondent reported lifetime nonmedical use of codeine, the code of 4 that was assigned to the variable in the initial processing was reassigned a code of 1 as part of the editing procedures. If the respondent did not choose codeine from the list of drugs in question PR04A, but reported nonmedical use of another pain reliever from the list, then the code of 98 in the variable for codeine (which had been assigned during the remapping of "enter all that apply" variables, as described in Section 2.2.2) was replaced with a code of 6 during the editing process.

Use of the code of 6 was intended to indicate to analysts that not choosing an available response from the list was not exactly the same as an answer of "no" in questions that required respondents explicitly to answer "yes" or "no" about a behavior of interest. In other words, a response of "no" in a "yes/no" type of question can be thought of as an *active* indication that the behavior or characteristic of interest did not apply, whereas not choosing a response in an "enter all that apply" list can be thought of as a *passive* indication that a particular behavior did not apply. In practice, however, not choosing a response from a list often was treated as being

equivalent to a response of "no" in subsequent editing steps. For example, see the discussion about editing for the most recent nonmedical use of the pain reliever OxyContin® in Section 6.2.2 in Chapter 6.

2.4.5 "Back-Editing" Based on "OTHER, Specify" Data

"Back-editing" refers to situations in which answers to a given question can be used to make inferences about how a preceding question *within the same module* should have been answered. Specifically, the principles and procedures that are discussed in this section refer to use of "OTHER, Specify" data to go "back" and edit an earlier variable according to what respondents specified in a later series of questions for that module.

In particular, a type of inconsistency that could commonly occur in the NSDUH data involved situations in which respondents did not answer a question affirmatively (e.g., in question LS01A, whether they ever used LSD). In the same module, however, they reported something in "OTHER, Specify" items that indicated that the other question should have been answered as "yes" (e.g., specifying use of LSD as some "other" hallucinogen that the respondent had ever used). When respondents specified something that corresponded to an item they had been asked about previously, but they had not answered that previous item as "yes," then the editing procedures assigned a value of "yes" to the relevant question. The following code typically was used when a response of "yes" was logically inferred: 3 = Yes LOGICALLY ASSIGNED.

If there was a lead to the "OTHER, Specify" question that was in the form of a "yes/no" question (e.g., "Have you ever, even once, used any other hallucinogens besides the ones that have been listed?"), the affirmative answer was retained in the lead to the "OTHER, Specify" question (i.e., having ever used any other hallucinogens), in addition to the inference being made that the answer to another question logically was "yes" (e.g., "Have you ever, even once, used LSD, also called 'acid'?"). The redundant specify code also was retained in the "OTHER, Specify" variable(s) to indicate to analysts the source of the logically inferred "yes" value.

This principle also applied to the editing of variables in "enter all that apply" questions based on answers in "OTHER, Specify" data. The following code typically was used when a response was logically inferred in a variable in an "enter all that apply" question: 3 = Response entered LOGICALLY ASSIGNED. For example, if a respondent did not choose codeine from the list in question PR04A, but specified codeine as another pain reliever that he or she misused, then it was inferred that codeine logically should have been chosen from the list. The individual edited variable corresponding to lifetime nonmedical use of codeine was assigned a code of 3.

2.4.6 "Flag and Impute" Principles for Drug Use Data

The editing procedures for establishing when respondents last used a substance are critical for creating final published estimates from NSDUH of the prevalence of substance use in the United States. In addition, data from core drug use modules on most recent use are important for establishing whether skipped questions in noncore modules truly were not applicable or if there might be some question about whether these skipped questions might have applied to the respondent. For example, respondents who reported that they used cocaine in their lifetime but

that they last used it more than 12 months ago were not asked questions about cocaine dependence or abuse. However, if any of these respondents also reported that they first used cocaine at their current age, these reports of recent initiation would suggest that they may have used cocaine in the past 12 months, in which case they should have answered the dependence or abuse questions for cocaine.

Under the deterministic edits for the old PAPI format, as a general rule, if a respondent indicated in one question on a substance's answer sheet that he or she had never used a substance and indicated use of that substance in another question on the answer sheet, logical editing coded the person as a user of that substance. If a respondent reported two (or more) different answers on the same answer sheet with respect to how recently he or she had used a substance, the editing procedures typically assigned the category indicating the more recent use. Relatively little statistical imputation was done to the PAPI recency variables following the editing step. A drawback of this approach was that decisions to infer more recent use could have an appreciable impact on estimates of use in the past 30 days for less commonly used substances, such as cocaine and heroin.

Conversion of the instrument to a CAI format in 1999 provided an excellent opportunity to reexamine the procedures and underlying assumptions for editing the recency variables. Further, the logic in the CAI instrument, in which respondents were skipped past questions that did not apply to them, precluded the same kinds of edits for the CAI recency variables as were done for PAPI.

Four possible ways of editing the data for most recent use of a drug were examined as part of the methodological research for data processing and estimation procedures using the CAI data (Kroutil & Myers, 2002). The flag and impute rule that was adopted for editing the CAI recency-of-use variables flags inconsistencies between a recency variable and related variables but does not make a decision about the final recency category. Rather, this rule leaves these inconsistent recency-of-use data to be statistically imputed.

For example, if a respondent originally reported last using a drug more than 12 months ago but also reported first using it at his or her current age, this procedure inferred that the respondent was *at least* a lifetime user. In the imputation procedures, this case's data for most recent use was imputed to indicate most recent use in any period (i.e., in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago). Also, the data on past year initiation that were inconsistent with most recent use of the drug more than 12 months ago were set to bad data (Section 2.4.3), which treated the initiation data as missing. Consequently, if the respondent was imputed to have last used the drug at some point in the past 12 months (including use in the past 30 days), then the respondent could be imputed to have initiated use of the drug within the past 12 months or more than 12 months ago based on data from a donor whose reports of first use and most recent use were consistent. However, if the respondent was imputed to have last used the drug more than 12 months ago, then initiation data from the donor respondent also would be consistent with initiation more than 12 months prior to the interview date.

The beauty of this edit rule lies in its simplicity: If a respondent gives an answer within a substance's module that conflicts with the original answer to the recency-of-use question, then

the recency variable is statistically imputed using data from a suitable donor record without these inconsistent data (see Chapter 3). A second attractive feature of this rule is that if the respondent provides conflicting information, it is not necessary to try to deduce from the data when the respondent last used the substance. Moreover, this rule does not automatically discount indications of more recent use than the respondent originally reported in a recency question, nor does it automatically infer that the respondent last used a substance more recently than he or she originally reported (as was the case with editing procedures for the former PAPI data). However, if final assignment occurred to indicate use in a more recent period than the respondent originally reported, this decision typically was made through statistical imputation rather than deterministic editing.[19]

[19] Limited exceptions that involved deterministic editing of recency variables are discussed in Chapter 6.

3. Imputation and the Predictive Mean Neighborhood Methodology

3.1 Introduction

As with most large-scale sample surveys, the respondent datasets for the 2012 National Survey on Drug Use and Health (NSDUH) contained missing responses for some items, inconsistent or invalid responses, and violations of skip patterns. Although the survey instrument was designed to enforce skip patterns and to perform some consistency checks as data were collected, invalid and inconsistent responses still occur. These response errors are a source of bias in the analysis of NSDUH data (Cox & Cohen, 1985).

Deterministic editing to correct erroneous and inconsistent responses and to replace missing values is appropriate when a unique association exists between predictor variables and the variable to be predicted (Cox & Cohen, 1985). For instance, gender often can be inferred from the respondent's relationship to the head of a household (e.g., son, daughter). However, even when good predictor variables are present, an unambiguous prediction may not be possible for every record having missing or faulty data (e.g., "cousin" does not clarify the gender of a respondent). In such cases, the remaining faulty or missing data often are replaced with statistically imputed data.

"Imputation" is the term used to describe the replacement of missing data with plausible values. Most commonly, imputation is used when a respondent answers some questions on a survey but not others. This is a condition known as "item nonresponse." By contrast, when a selected individual does not respond to any question on the survey at all, or does not respond to enough key questions for the case to be useful for research purposes, this is a condition referred to as "unit nonresponse." In such cases, weighting adjustments are normally employed to account for these missing data. As an initial step, prior to any processing of the data, unit nonrespondents were discarded, and only unit respondents (i.e., item respondents and item nonrespondents for any given questionnaire item) were included in the subsequent editing, imputation, and analysis of NSDUH data.

Once processed, imputed values cannot be distinguished from nonmissing values for a given variable in the final dataset. Therefore, observations with imputed data must be identified with a concomitant indicator variable. The vast majority of imputation-revised variables for the 2012 NSDUH have the prefix "IR" attached to their names.[20] Although no missing data were possible for gender because a response to this item was required before the interview could proceed, the "IR" prefix for IRSEX was maintained for continuity with past years. Each imputed variable has an associated indicator variable, identified by the prefix "II" that can be used to identify which values were imputed and which were not. For some imputation-revised variables, additional imputation indicators were created with the prefix "II2." These indicators gave more

[20] Exceptions to this rule included the imputation-revised employment status variables EMPSTAT4 and EMPSTATY and the core-plus-noncore methamphetamine and stimulant variables CPNMTHFG, CPNMTHYR, CPNMTHMN, CPNSTMFG, CPNSTMYR, and CPNSTMMN.

details about the source of the imputed or logically assigned value. The levels of a typical "II" imputation indicator are as follows:

1 = From questionnaire

2 = Logically assigned

3 = Statistically imputed

9 = Legitimate skip

Assignment of a code of 9 to an imputation indicator reflected the prior assignment of a legitimate skip code as part of the editing process (Section 2.4.2). Also, if a question had missing or ambiguous data and this question governed the skip logic for subsequent questions, respondents typically were skipped out of the subsequent questions. Respondents requiring imputation for the variables that governed a skip pattern typically receive a value of 3 (statistically imputed) for the imputation indicators associated with both the governing variables and the imputed variables that were nested within the skip pattern. For example, if a respondent had missing data for whether he or she had ever used a particular substance and was imputed to have never used it, the imputation indicators for recency of use of the substance, age at first use, frequency of use in the past 12 months (if applicable), and frequency of use in the past 30 days (if applicable) all were assigned a code of 3.

3.2 Development of the Predictive Mean Neighborhood Methodology

Various methods of imputation have been used since the NSDUH was first administered in the early 1970s.[21] With the expansion of the NSDUH sample size in 1999, the predictive mean neighborhood (PMN) method for imputation was implemented and is currently used for most variables. PMN is designed to incorporate the complex interrelationships among items in the current NSDUH, thus maintaining data consistency within individual respondent records. Table 3.1 provides a summary of the types of imputation procedures used for each of the variables imputed in the NSDUH samples from 1999 through 2012.

Table 3.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year

Variable	1999[1]	2000	2001	2002-2003	2004-2012
Interview Date	Random[2]	Random	None	None	None
Age	None[3]	None	None	None	None
Birth Date	None	Random	Random	Random	Random
Gender	None	None	None	None	None
Race	USHD[4]	PMN	PMN	PMN	PMN
Hispanic or Latino Origin Indicator	USHD	PMN	PMN	PMN	PMN
Marital Status	USHD	PMN	PMN	PMN	PMN
Hispanic or Latino Origin Group	USHD	PMN	PMN	PMN	PMN
Education	USHD	USHD	PMN	PMN	PMN

[21] Prior to the 2002 survey year, when it was renamed, the National Survey on Drug Use and Health (NSDUH) was originally known as the National Household Survey on Drug Abuse (NHSDA).

Table 3.1 Summary of Item Imputation Procedure Used, by Variable and NSDUH Year (continued)

Variable	1999[1]	2000	2001	2002-2003	2004-2012
Employment Status	USHD	USHD	PMN	PMN	PMN
Immigrant	Not imputed	Not imputed	Not imputed	WSHD[5]	PMN
Health Insurance	PMN	PMN	PMN	PMN[6]	PMN
Lifetime Drug Usage	PMN	PMN	PMN	PMN	PMN
Recency and Frequency of Use[7]	PMN	PMN	PMN	PMN	PMN
Age at First Use	PMN	PMN	PMN	PMN	PMN
Age at First Daily Cigarette Use	PMN	PMN	PMN	PMN	PMN
Personal and Family Income (Binary)	PMN	PMN	PMN	PMN	PMN
Personal and Family Income (Finer Categories)	PMN	PMN	PMN	PMN	PMN
Nicotine Dependence	Not imputed	Not imputed	Regression	Regression	Regression
Household Size (Roster-Derived)	PMN	PMN	PMN	PMN	PMN
Other Household Composition (Roster-Derived)	PMN	PMN	PMN	PMN	PMN

[1] The 1999 survey year also included a paper-and-pencil interviewing sample. The procedures listed here are from the computer-assisted interviewing sample.
[2] "Random" refers to a random assignment within a quarter for the interview date and a random assignment using age and interview date for the birth date.
[3] "None" means that no missing values were encountered after editing, and thus no imputation was necessary. For gender (from the 2002 survey onward) and age, missing values were precluded by design (see Chapter 4).
[4] "USHD" refers to the unweighted sequential hot-deck method of item imputation described in this report (see Section 3.2.1.1).
[5] "WSHD" refers to the weighted sequential hot-deck method of item imputation described in this report (see Section 3.2.1.2).
[6] Although PMN was the method used for health insurance in all years since the 1999 survey, imputation also was applied to more detailed health insurance variables in the surveys from 2002 onward.
[7] "Recency and Frequency of Use" included variables measuring recency of use, 12-month frequency of use, 30-day frequency of use, and binge drinking frequency in past 30 days. "Binge drinking" was defined as having five or more drinks on the same occasion on a given day.

3.2.1 Previously Used Hot-Deck Imputation Methods

With any method of imputation, missing responses for a particular variable (hereafter, termed "base" variable) are replaced by values from similar respondents with respect to a number of characteristics (hereafter, "auxiliary variables"). If "similarity" is defined in terms of a single predicted value from a model, these auxiliary variables can be represented by that value. The respondent with the missing value for the base variable is called the "recipient," and the respondent from whom values are borrowed to replace the recipient's missing value is called the "donor." Donors and recipients are distinguished by the completeness of their records with regard to the variable(s) of interest (i.e., the donor has complete data for that variable, and the recipient does not). The term "hot deck" is used to refer to imputations made on recipient base variables using donor values from the same dataset. The PMN methodology utilized on the NSDUH is a specialized hot-deck method and is described in greater detail later in this chapter. For more information on the general hot-deck method of item imputation, see Little and Rubin (1987, pp. 62-67).

For the 2012 NSDUH, the only imputations that did not incorporate the PMN method were those used for the birth date, date of first use, and nicotine dependence variables, described in Section 4.2.5, Section 6.3.3.4, and Chapter 7, respectively. Two other hot-deck methods—unweighted sequential hot deck (USHD) and weighted sequential hot deck (WSHD) (Cox, 1980, pp. 721-725; Iannacchione, 1982)—were used in past surveys.[22] In the sections that follow, the features and limitations of USHD, WSHD, and the random nearest neighbor hot deck (NNHD) are discussed as background for the development of the PMN methodology.

3.2.1.1 Unweighted Sequential Hot Deck

In a sequential hot-deck procedure, data are first ordered using specific criteria, and the last reported value in the sequence is substituted for each missing value as the data are processed. In USHD, the selection of a response for imputation purposes is independent of the sampling weight associated with the data record from which the response is taken and the data record to which a response is being imputed. USHD imputation is, therefore, based upon the tacit assumption that nonrespondents would answer in a manner similar to that of respondents immediately adjacent to them in an appropriately sorted data file and hence that the data associated with the nearest neighbor are appropriate for the imputation of missing values (Cox, 1980, p.721).

Implementation of the USHD method (and of hot-deck methods, in general) involves three basic steps:

1. **Construct imputation classes**. When there is a strong logical association between the base variable and certain auxiliary variables, the dataset is partitioned by these auxiliary variables, and imputation procedures are implemented independently within the resulting imputation classes defined by the cross of these auxiliary variables.
2. **Sort the analytic file.** Within each imputation class, the file is sorted by auxiliary variables relevant to the item being imputed. The sort order of the auxiliary variables is chosen to reflect the degree of importance of the auxiliary variables in their relation to the base variable being imputed (i.e., those auxiliary variables that are better predictors for the item being imputed are used as the first sorting variables). In general, two types of sorting procedures—a straight sort and a serpentine sort[23]—were used in previous surveys to sort the files prior to imputation.
3. **Replace missing values with imputed values.** The sorted file is read sequentially. Each time an item respondent is encountered (i.e., the base variable is nonmissing),

[22] The USHD method was used exclusively for the 1991-1998 surveys, for the paper-and-pencil interviewing sample from the 1999 survey, and for all demographic variables in the computer-assisted interviewing sample from the 1999 survey. In the 2002-2003 surveys, missing values in the immigrant variables required WSHD imputation. Note, however, that the USHD and WSHD methods have not been used on the NSDUH since the 2000 and 2003 survey years, respectively.

[23] Under a straight sort, a set of variables is sorted in ascending order by the first variable specified. Then, within each level of the first variable, the file is sorted in ascending order by the second variable specified, and so forth. In a serpentine sort, a set of variables is sorted so that the direction of the sort (ascending or descending) for subsequent variables changes each time the value of the preceding variable changes. The serpentine sort has the advantage of minimizing the change in the entire set of auxiliary variables every time any one of the variables changes its value. For an example of each, see Appendix A of the 2009 imputation report (Ault et al., 2011).

the base variable response is stored, updating the donor response. Any subsequent nonrespondent in the file receives the stored donor response, which in turn results in a statistically imputed response. Because the file is sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matches the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

For any particular item being imputed under USHD, there is the risk of several nonrespondents appearing next to one another on the sorted file; in this situation, each would receive imputed values from the same donor. To detect this problem on the NSDUH in the survey years prior to 2001, the imputation donor was identified for every item being imputed, and frequencies by donor were examined. If several nonrespondents were aligned next to one another after sorting, sort variables were added or eliminated, or the ordering of the sort variables was modified, to ensure that multiple nonrespondents did not comprise adjacent records on the resulting file.

3.2.1.2 Weighted Sequential Hot Deck

WSHD improves upon USHD by incorporating the sampling weights when replacing missing values among recipient records in the final hot-deck assignment step. The earlier steps taken to impute for missing values under the WSHD method are the same as those for the USHD method; as in USHD, WSHD requires the formation of imputation classes and appropriate sorting (straight or serpentine) of the analytical file.

The WSHD procedure used in surveys prior to 2004 followed directly from Cox (1980). Specifically, once the imputation classes were formed, the data were divided into two datasets: one for respondents and one for nonrespondents. Scaled weights v_j were then derived for all nonrespondents using the following formula:

$$v_j = w_j s_+ / w_+ ; j = 1, 2, \ldots, n,$$

where n is the number of nonrespondents, w_j is the sample weight for the jth nonrespondent, w_+ is the sum of the sample weights for all the nonrespondents, and s_+ is the sum of the sample weights for all the respondents. The respondent data file was partitioned into zones of width v_j, where the imputed value for the jth nonrespondent was selected from a respondent in the corresponding zone of the respondent data file. This selection algorithm is an adaptation of Chromy's (1979) sequential sample selection method.

WSHD controls the number of times a donor can be selected and allows each respondent the chance to be a donor because a respondent is selected within each v_j. Consequently, the most important benefit of the weighted sequential hot-deck method is the elimination of bias in the estimates of means and totals, particularly when the response rate is low or when the covariates explain only a small amount of variation in the specified variable. In addition, many surveys sample subpopulations at different rates, and using the sample weights allows the imputed data for the nonrespondents to have the same mean (for the specified variables) as the respondents. In other words, the weighted hot deck preserves the respondent's weighted distribution in the imputed data (Cox, 1980).

3.2.1.3 Unweighted Random Nearest Neighbor Hot Deck

Another commonly used imputation method—one not directly used on the NSDUH, but related to the PMN method—is random nearest neighbor hot deck (NNHD) (Little & Rubin, 1987, p. 65). With this method, a donor set or neighborhood deemed "close to" the recipient, with respect to a number of covariates, is used to select a donor at random. The distance between the values of the recipient and potential donors for each of the auxiliary variables is calculated, and then the donors for the neighborhood are chosen such that the maximum of these distances is less than a certain threshold value, referred to as "delta." This neighborhood is restricted, using imputation classes described previously, so that the potential donors' values of the base variable are consistent with the recipient's preexisting nonmissing values of related variables.

Because a distance function is used to define "closeness" between the recipient and a donor under NNHD, there is less of a problem of sparseness of the donor class when imputing for continuous variables. It should be noted, however, that the distance function involving categorical or nominal variables is typically ad hoc and often hard to justify.

3.2.2 Advantages of the Predictive Mean Neighborhood Methodology

The PMN methodology developed for and implemented on the 1999 NSDUH was an attempt to address the shortcomings, while retaining the positive characteristics, of the hot-deck imputation methods discussed above. It is a combination of two commonly used imputation methods: non-model-based NNHD (Little & Rubin, 1987, p. 65) and a modification of Rubin's model-assisted predictive mean matching (PMM) method (Rubin, 1986). The PMN method enhances Rubin's PMM method, in that PMN can be applied to both discrete and continuous variables, either individually or jointly. PMN also enhances the NNHD method for discrete variables so that the distance function used to find neighbors is no longer in terms of the original predictor variables and therefore does not require arbitrary scaling.

In addition, the PMN methodology offers the following advantages over the imputation methods employed on earlier NSDUHs:

- **A greater number of auxiliary variables may be used to determine donors.** Using a model-based hot-deck technique like PMN allows auxiliary variables to be incorporated in two ways: first, as covariates in models, and second, in likeness constraints[24] applied to potential donors. Under USHD and WSHD, the number of auxiliary variables is limited in part by the problem of sparse neighborhoods; donors must match recipients for all variables used to form imputation classes. If too many variables are used to form imputation classes, some classes may be very small and contain few or no item respondents to serve as donors. By contrast, under PMN, the donors need only be "close" to the recipients with respect to the predicted values determined by the models, even when the models include numerous covariates. Moreover, PMN ensures that a sufficient number of potential donors comprise the donor neighborhood, so that likeness constraints may be applied on the donor set as needed.

[24] Likeness constraints are flexible constraints that govern the similarity between donors and recipients. See Section 3.3.1.3 for details.

- **Relative importance of auxiliary variables is determined by standard estimating equation techniques.** Under USHD and WSHD, as implemented, the selection of classing and sorting variables was sometimes ad hoc, and in the former instance, weights were not utilized. In PMN, by contrast, objective criteria based on a more rigorous methodology (i.e., regression) quantify the relationship between a given covariate and the response variable in the presence of other covariates, so that the response variable itself is indirectly used to determine donors. Further, the sampling weights can be incorporated in PMN regression models without difficulty.
- **Internal consistency of the post-imputation record is guaranteed.** In PMN, the donor pool can be restricted to those making the post-imputation record logically consistent. For example, if a recipient must receive a cocaine past year frequency of use between 30 and 50, the donors can be restricted to ensure that the recipient receives such a value. In USHD and WSHD, the classing and sorting variables cannot easily be used to guarantee this; there may not be a donor in the imputation class who will create a consistent record.
- **Correlations across response variables are accounted for by making the imputation multivariate.** In comparison with other model-based methods, discrete and continuous variables can be handled jointly and relatively easily in PMN by using the idea of sequential univariate modeling. Further, differential weighting factors can be objectively assigned to different elements of the predictive mean vector depending on the variability of predicted means in the dataset.

3.3 Implementation for the Predictive Mean Neighborhood Methodology

The implementation of PMN on the NSDUH involves three basic steps: response propensity (RP) adjustment, prediction (PRD) modeling, and hot-deck imputation. At the most basic level, the RP adjustment reallocates the weights of the item nonrespondents to item respondents; the prediction model calculates predicted means for both; and the hot-deck step assigns final values to the item nonrespondents based on a distance function derived from these predicted means. These steps are described in more detail in the following sections and are combined in three different ways, called PMN "types" (Section 3.4), to complete imputation procedures.

3.3.1 Step 1: Response Propensity Adjustment

Response propensity is defined as the probability of response, whether at the unit level or item level. The purpose of response propensity is to adjust the sampling weights for item nonresponse so that the item respondent weights that are used only during the imputation process are representative of the entire domain of interest. In the response propensity step of PMN, the item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special case of the generalized exponential model (GEM)[25] developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

[25] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of Chen et al. (2014).

There are four key inputs to the item response propensity modeling step:

1. **Analysis weights.** For all imputation procedures, the best available analysis weight is used as an input to the NSDUH imputation procedures. Because of the timing of 12-month processing and, in particular, the coordination between the weighting and imputation tasks in each NSDUH year, most variables that undergo imputation utilize the preliminary analysis weight, PANALWT. For those variables that are processed later in the annual cycle, the final analysis weight ANALWT may be used instead, if it is available at the time of imputation processing for that variable. The pair variables described in Chapter 11 utilize yet another weight, PAIRWT. See Chen et al. (2014) for full descriptions of the preliminary and final analysis weights. See Westlake, Chen, and Gordek (2014) for full descriptions of the pair weights.
2. **Domain indicator.** In this report, a "domain" is defined as the set of respondents who are included in models and for whom predicted means are calculated. For many NSDUH variables that undergo imputation, the domain includes all unit respondents. For others, the domain is a subset of unit respondents. For example, the marital status question is only asked of respondents aged 15 or older. In this case, the domain indicator is set to zero for respondents aged 12 to 14 and to one for respondents aged 15 or older. The domain indicator is an important input to the tables in Appendix A, where item response rates are reported for each variable that undergoes imputation on the NSDUH.
3. **Item response indicator.** The item response indicator is set to zero for item nonrespondents and to one for item respondents. GEM uses this indicator to identify the item nonrespondents and item respondents when reallocating the weights appropriately. The item response indicator is an important input to the tables in Appendix A, where item response rates are reported for each variable that undergoes imputation on the NSDUH.
4. **Covariates.** GEM uses a predetermined list of covariates to allocate the weights from item nonrespondents to item respondents. The covariates tend to be variables that are correlated with (1) the propensity to respond, (2) the variable that is undergoing imputation, or (3) both. The goal is to avoid bias in the prediction models by allocating the weights of the item nonrespondents to similar item respondents, without too greatly inflating the variance of the estimates that utilize these weights (Chen et al. 2014). Appendix D lists the starting and final covariate lists for each response propensity model fit in NSDUH imputation procedures.

3.3.2 Step 2: Prediction Modeling

Utilizing the response propensity-adjusted weights that were derived in the previous step, the prediction model calculates predicted means, which are used in the hot-deck step(s) to create neighborhoods and select donors. The dependent variable in the model is usually the variable, or some transformation of that variable, that is undergoing imputation. Each model is built using only those cases within the domain with complete responses for that item. Predicted means are then calculated for all of the domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, logistic regression models are used for the prediction models. For continuous variables, linear regression models are fit. For count variables, Poisson regression models are used. For response variables that are proportions (e.g., months on welfare, see Chapter 9), a logit transformation is applied to the proportion, and a linear regression model is utilized. The variable sets in which some transformations of the response variables were implemented include the noncore demographics (Chapter 5), drugs (Chapter 6), and income (Chapter 9).

The goal of any prediction model is good prediction, so these models tend to start with long lists of covariates. Appendix D lists the starting and final covariate lists for each prediction model fit in NSDUH imputation procedures. In contrast to explanatory (association) models where model parsimony is a relevant metric of a model's appropriateness, the focus in a prediction setting is on the predicted values only.

The SUDAAN software package is used to fit nearly all the prediction models used in the NSDUH.[26] All covariates from the applicable starter list are utilized unless SUDAAN produces warning messages, which indicate nonconvergence or model instability. In these cases, the standard errors of the regression coefficients are used to make decisions about which covariates to drop from the models; covariates are dropped until SUDAAN no longer produces these warning messages. The primary advantage of using SUDAAN to fit prediction models is that the standard errors associated with the regression coefficients properly account for the complex survey design. The predicted means are the same using SUDAAN as they are using, for example, the analogous SAS procedure (given the same set of covariates), but the decision on which covariates to drop in the event of model instability or nonconvergence is more informed under SUDAAN because the standard errors are more accurate.

In the particular case of some of the logistic regression models, the warning messages produced by SUDAAN may be triggered when a cross-classification of the outcome variable and a covariate has empty or nearly empty cells. Covariates of this type are highly correlated with the outcome variable but cannot be used in the prediction model. However, they are often used in the hot-deck step to identify suitable donors.

For the types of regression-based prediction models used for each variable that underwent imputation using PMN, see Table 3.2.

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Demographics			
Marital Status	15 years or older	Multinomial Logistic	MULTILOG
Race	All respondents	Multinomial Logistic	MULTILOG
Hispanic or Latino Indicator	All respondents	Binomial Logistic	RLOGIST
Hispanic or Latino Group	Hispanics	Multinomial Logistic	MULTILOG
Education Level	All respondents	Multinomial Logistic	MULTILOG

[26] The two exceptions are the finer income categories, described in detail in Chapter 9, and the "old method" health insurance, described in detail in Chapter 10.

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Employment Status	15 years or older	Multinomial Logistic	MULTILOG
Immigrant Status: Born-in-U.S. Indicator	All respondents	Binomial Logistic	RLOGIST
Immigrant Status: Age of Entry	Not born in U.S.	Simple Linear	REGRESS
Drugs			
Lifetime Drug Use	All respondents	Binomial Logistic	RLOGIST
Recency of Drug Use, "Hierarchical" Drugs	All lifetime users for past year vs. not past year; all past year users for past month vs. not past month	Binomial Logistic	RLOGIST
Recency of Drug Use, Pipes	All lifetime users	Binomial Logistic	RLOGIST
Recency of Drug Use, All Other Drugs	All lifetime users	Multinomial Logistic	MULTILOG
12-Month Frequency of Drug Use	All past year users	Simple Linear	REGRESS
Daily Drug Use over Past 30 Days, Cigarettes, Chewing Tobacco, and Snuff	All past month users	Binomial Logistic	RLOGIST
30-Day Frequency of Drug Use, Cigarettes, Chewing Tobacco, and Snuff	All past month users except those who used daily over the past 30 days	Simple Linear	REGRESS
30-Day Frequency of Drug Use, All Other Drugs	All past month users	Simple Linear	REGRESS
Age at First Drug Use	All lifetime users	Simple Linear	REGRESS
Household Composition			
Total Number of Rostered Persons	All respondents	Poisson	LOGLINK
Total Number of Children Younger than 18	All respondents	Poisson	LOGLINK
Total Number of Persons Aged 65 or Older	All respondents	Poisson	LOGLINK
Indicator of Whether the Respondent Has Family Members in Household	All respondents	Binomial Logistic	RLOGIST
Total Number of Respondent's Family Members in the Household (Excludes Foster Relationships)	All respondents	Poisson	LOGLINK
Total Number of Respondent's Family Members in the Household Younger than 18 (Excludes Foster Relationships)	All respondents	Poisson	LOGLINK

Table 3.2 Regression Models Used for Each Variable Imputed with Predictive Mean Neighborhood (continued)

Variable	Domain	Type of Regression Model	SAS/SUDAAN Procedure[1,2]
Total Number of Respondent's Family Members in the Household (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Total Number of Respondent's Family Members in the Household Younger than 18 (Includes Foster Relationships)	All respondents	Poisson	LOGLINK
Income			
Source of Income	All respondents	Binomial Logistic	RLOGIST
Months on Welfare	All respondents who received welfare payments or welfare services in the past year	Simple Linear	REGRESS
Total Income (Binary)	All respondents	Binomial Logistic	RLOGIST
Finer Category Income	All respondents	Time-to-Event (Survival)	LIFEREG
Health Insurance			
Health Insurance (Old Method)	All respondents	Binomial Logistic	LOGISTIC
Health Insurance (Constituent Variables Method)	All respondents	Binomial Logistic	RLOGIST

[1] SAS® software is a registered trademark of SAS Institute Inc. SUDAAN® is a registered trademark of Research Triangle Institute.

[2] See RTI International (2008) for more information on all SAS-callable SUDAAN procedures in this table except LIFEREG and LOGISTIC. See SAS Institute Inc. (2013) for more information on the LIFEREG and LOGISTIC procedures.

3.3.3 Step 3: Hot-Deck Imputation

After sampling weights have been appropriately adjusted in the response propensity step and predicted means have been calculated in the prediction step, the hot-deck step[27] of PMN is applied to select a donor for each item nonrespondent. The algorithm used to select donors is graphically displayed in the flow chart in Figure 3.1. Briefly, likeness constraints are loosened in an iterative fashion until PMN yields a nonempty donor neighborhood. Mahalanobis distance is then used to rank donors by closeness to the item nonrespondent, and a final donor is selected at random from a minimum of 30 candidate donors to supply imputed value(s) for a given recipient. Many of the hot-deck components used in PMN are described below and appear explicitly in the tables of Appendix E.

[27] There is one situation on the NSDUH in which the imputation is not a hot-deck step, but is a stochastic imputation based solely on the predicted mean(s) of the recipient, of the type described in Section 5.1 of the PMN imputation evaluation report by Ault et al. (in press). These ideas have their origin in Singh, Grau, and Folsom (2004), where Centered PMN is discussed as an alternative to PMN. This one exception is the provisional hot-deck step for imputation set 2 for the health insurance variables (Section 10.3.2.3).

Figure 3.1 Donor Selection Algorithm

3.3.3.1 Logical and Likeness Constraints

Logical constraints and likeness constraints are restrictions placed on the set of donors to make imputed values consistent with preexisting, nonmissing values of the item nonrespondents (recipients) and to make candidate donors as much like the recipients as possible. Logical constraints are fixed constraints that prevent logical inconsistencies between variables, and likeness constraints are flexible constraints that govern the similarity between donors and recipients.

The logical constraints are never removed, because to do so would risk the selection of a donor that produces an inconsistent post-imputation record. For example, for the employment status variable, if the item nonrespondent is known to be employed, but full-time vs. part-time status is unknown, the imputed value must come from a donor who is employed as well.

Likeness constraints are placed on the pool of donors to make the attributes of the neighborhood as close as possible to those of the recipient. For example, age and employment status are correlated. A likeness constraint exploits this correlation by requiring the donor's age to be within 5 years of the item nonrespondent's age, but likeness constraints may be loosened if they happen to force the donor pool to be empty.

One likeness constraint that is used in all hot-deck steps, regardless of the variable being imputed, is the delta constraint. This particular likeness constraint requires the donor's predicted mean to be within 5 percent (delta) of the item nonrespondent's predicted mean for each element of the predictive mean vector. If the predicted means are probabilities, the values of delta vary depending upon the value of the predicted mean.

Each delta is defined as 5 percent of the predicted probability if the probability were less than 0.5 and is defined as 5 percent of 1 minus the predicted probability if the probability were greater than 0.5. This allows for a looser delta for predicted probabilities close to 0.5 and a tighter delta for predicted probabilities close to 0 or 1. The range of values for delta across various predicted probabilities is shown in Table 3.3.

Table 3.3 Values of Delta for Various Predicted Probabilities

Predicted Probability (p)	Delta
$p \leq 0.5$	$0.05p$
$p > 0.5$	$0.05(1 - p)$

Logical constraints and likeness constraints, including the order in which likeness constraints are loosened for some variables, are presented in the tables in Appendix E.

3.3.3.2 Predictive Mean Vector

The predicted means from the prediction step play a central role in the donor selection algorithm depicted in Figure 3.1, through the construct of the predictive mean vector. The predictive mean vector is essentially a list of predicted means from the prediction modeling step.

In simple cases, the predictive mean vector contains only one element, such as the predicted age at which a respondent began using a drug. In complex cases, the predictive mean vector includes several elements from several different prediction models, such as the predicted recency and predicted frequency of use for a given drug.

When the prediction model is a logistic regression model, predicted means are calculated for each level of the outcome variable. For example, the employment status variable that undergoes imputation has four levels: employed full time, employed part time, unemployed, or other. Therefore, a single prediction model is fit using a four-level outcome variable, yielding predicted probabilities for each level, as follows:

- E1: P(respondent is employed full time)
- E2: P(respondent is employed part time)
- E3: P(respondent is unemployed)
- E4: P(other)

Note, however, that the predictive mean vector for the employment status variable contains only three elements. It does not include the predicted probability for the reference cell, which in this case is the "other" level, since that level is implicitly defined by the presence of the other three predicted means.

Occasionally, the predicted means are adjusted so that they are made conditional on what is known for a given respondent. Continuing the example above, some respondents report that they have a job but are unclear about the number of hours they usually work in a week. Because the NSDUH definition of part-time versus full-time employment status was based on working a minimum of 35 hours in a usual week, the predictive mean vector is made conditional on employment of any sort for these respondents. Therefore, the single predicted mean used for these respondents is equal to E1/(E1 + E2), P(respondent is employed full time | respondent is employed). Conditional probabilities are also used in the binary income hot-deck step and the drug recency/frequency hot-deck steps.

Predictive mean vectors are presented in the tables in Appendix E.

3.3.3.3 Univariate vs. Multivariate Matching and Assignment

If the predictive mean vector consists of only one element, *univariate* matching is used to select a donor. If the predictive mean vector consists of more than one element, *multivariate* matching is used to select a donor. The donor may also give values to the item nonrespondent for more than one variable, a situation known as multivariate assignment. Similarly, if the donor provides values for only one variable, the hot-deck step uses univariate assignment. Table 3.4 shows examples of NSDUH variables that were imputed using each of the four combinations of univariate/multivariate matching and assignment.

Table 3.4 Examples of Variables Imputed Using Each of the Four Combinations of Univariate/Multivariate Matching and Assignment

	Variables Imputed One at a Time (Univariate Assignment)	Variables Imputed in a Set (Multivariate Assignment)
Predictive Mean Vector Has One Element (Univariate Matching)	Hispanic/Latino Origin (Section 4.3.3)	Finer Category Income (Section 9.3.2)
Predictive Mean Vector Has More Than One Element (Multivariate Matching)	Marital Status (Section 4.3.1)	Lifetime Drug Use (Section 6.3.1)

Whether the hot-deck step employs univariate or multivariate matching, Mahalanobis distance is used to rank the donors by closeness to the item nonrespondent. The Mahalanobis distance is used instead of Euclidean distance in order to standardize the distance in terms of the population variances and covariances of vector components. It is given by

$$\sqrt{(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_R - \boldsymbol{\mu}_{NR})},$$

where $\boldsymbol{\mu}_R$ refers to the predictive mean vector for a given item respondent, and $\boldsymbol{\mu}_{NR}$ is the predictive mean vector for a given item nonrespondent. The matrix $\boldsymbol{\Sigma}$ is the variance-covariance matrix of the predictive mean vector, using the set of item respondents that comprise that domain. Because the square of the Mahalanobis distance is a monotone function of the distance itself, and only the ranking of the donors (instead of the absolute distance measure) is used in the algorithm, the additional step of taking the square root of the squared distance is not performed in practice.

3.3.3.4 Missingness Patterns

For many variables imputed on the NSDUH, item nonrespondents were segregated into patterns of nonresponse called missingness patterns. Missingness patterns arise in two ways. First, for sets of variables that underwent multivariate assignment, item nonrespondents were segregated into missingness patterns based on which variables were missing. Second, a new missingness pattern could emerge when logical editing restricted an item nonrespondent to only a subset of the variable's possible values. The example for employment status discussed above applies here as well: respondents whose employment status was completely unknown had a different missingness pattern than did those who were known to be employed. Often, different predictive mean vectors were used, and different constraints were applied, for different missingness patterns. Many of the tables in Appendix E are segregated by missingness pattern for this reason.

3.3.3.5 Final Assignment of Donor Values

Logical and likeness constraints are used to form a neighborhood of potential donors from the pool of item respondents within each missingness pattern. Logical constraints are

always imposed to maintain internal consistency, whereas likeness constraints are removed or relaxed in a predetermined order until this donor neighborhood is nonempty. Once a nonempty neighborhood is found, the rest of the PMN donor selection algorithm depends on whether or not the delta constraint was applied.

If the delta constraint was applied, all the members of the neighborhood are similar to the recipient with respect to the predictive mean vector. The final donor is then randomly selected with equal probability from among the "closest" (in terms of Mahalanobis distance) 30 members of the neighborhood; potential donors whose Mahalanobis distance from the recipient are equal ("ties") are accounted for in the donor selection algorithm depicted in Figure 3.1. If, on the other hand, the delta constraint was not applied, to ensure that the final donor is as close to the item nonrespondent as possible with respect to the predicted means, the donor with the smallest Mahalanobis distance is selected as the final donor. If there is more than one "closest" donor (i.e., there are ties), the final donor is randomly selected with equal probability from among the closest donors. At the conclusion of the hot-deck step, the item nonrespondent receives values from the selected donor for a single variable (in the univariate assignment case) or for a set of variables (multivariate assignment).

3.4 Predictive Mean Neighborhood "Types"

There are three types of PMN as applied on the NSDUH: Type 1, *single response propensity (RP)/single prediction (PRD)* (Section 3.4.1); Type 2, *multiple RP/multiple PRD* (Section 3.4.2); and Type 3, *single RP/multiple PRD* (Section 3.4.3). Each of the three PMN types is a coordinated application of the three basic steps of PMN discussed in Section 3.3.

In PMN, an imputation "set" is a set of variables for which a single donor is used in the final hot-deck step.[28] Sets are formed based on the extent of correlation among variables and the level of missingness in the data. Variables with few missing values and no strong relationships with other variables tend to be processed in an imputation set by themselves. Closely related variables tend to be processed together in the same set to preserve, as much as possible, correlations between variables in the data. However, the more variables that are included in a multivariate set, the less likely it is that a nonempty neighborhood can be found using the delta constraint. Even though there are many advantages to using a multivariate imputation set, one disadvantage in several instances is not being able to apply the delta constraint.

Table 3.5 lists the imputation sets for each variable group discussed in this report and the PMN type used to process each set.

[28] Section 3.4.2 defines and discusses the differences between *provisional* and *final* hot-deck steps in the context of PMN.

Table 3.5 PMN Types Applied to Each Variable Group and Imputation Set

Variable Group	Imputation Set	PMN Type
Core Demographics	All (5 sets)	Type 1 (Single RP/Single PRD)
Noncore Demographics	All (3)	Type 1
Drugs	Lifetime	Type 3 (Single RP/Multiple PRD)
	Recency of Pipe Use	Type 1
	Recency/Frequency, other drugs (13)	Type 2 (Multiple RP/Multiple PRD)
	Cigarette Ever Daily Used	Type 1
	Age at First Use (14)	Type 1
Roster	All (8)	Type 1
Income	Binary	Type 3
	Finer Categories	Type 1
Health Insurance	Old Method	Type 3
	Constituent Variables Method, Stage 1	Type 3
	Constituent Variables Method, Stage 2	Type 1
Roster Pair	Pair Relationship	Type 1
	Multiplicities (6)	Type 1
	Household Counts, Sibling-Sibling and Spouse-Spouse (4)	Type 1
	Household Counts, Parent-Child	Type 2

PRD = prediction; PMN = predictive mean neighborhood; RP = response propensity.

3.4.1 Type 1: Single Response Propensity/Single Prediction

PMN Type 1, the single RP/single PRD type, involves a single iteration of the three basic steps described in Section 3.3: response propensity, prediction, and hot-deck imputation. Many variables that undergo imputation in the standard processing cycle use this type, including all the demographics and roster variables and the age-at-first-use drug variables. Figure 3.2 illustrates the single RP/single PRD type of PMN imputation.

Figure 3.2 PMN Type 1: Single Response Propensity/Single Prediction

Usually the single RP/single PRD type involves univariate assignment in the hot-deck step,[29] but it may involve univariate or multivariate matching, depending on the prediction model. If the prediction model is a dichotomous logistic regression, linear regression, or Poisson regression model, univariate matching is used because the model produces only one predicted mean. If, on the other hand, the prediction model is a polytomous logistic regression model, multivariate matching is used because the model produces more than one predicted mean (i.e.,

[29] Finer income categories is an example of an imputation set that uses the single RP/single PRD type, but its hot-deck step utilizes multivariate assignment. If the item nonrespondent is missing the finer income category at both the personal and family level, the donor will provide values for both variables in a single hot-deck step. The prediction model is fit using the family-level finer income category.

the predicted probability associated with each level of the response variable). In either implementation, there is only one prediction model.

In the single RP/single PRD type, for the univariate assignment case, the item response indicator is based on the single variable that is being assigned in the hot-deck step. If the single variable is missing, the case is an item nonrespondent; otherwise, the case is an item respondent. In the multivariate assignment case, the case is an item respondent if all variables that are assigned in the hot-deck step are nonmissing.

3.4.2 Type 2: Multiple Response Propensity/Multiple Prediction

PMN Type 2, multiple RP/multiple PRD, involves multiple iterations of the single RP/single PRD type. However, for all iterations except the last, the hot-deck step is provisional instead of final and involves univariate matching and univariate assignment.[30] These provisional hot-deck steps tend to be straightforward with respect to constraints and predictive mean vectors, because their only purpose is to fill in missing values so that variables earlier in the sequence can be used as covariates in the RP and PRD models for variables later in the sequence.[31] In the last iteration, a final hot-deck step is completed, where final imputed values are assigned for all variables involved in the models. This final hot-deck step always involves multivariate matching and multivariate assignment. The predicted means from all PRD models are used in this final hot-deck step, and a single record is used to fill all the missing values, thus preserving the relationships among the variables of interest. This is the most refined type of PMN. The recency and frequency variables (within each drug family) follow this type. Figure 3.3 illustrates the multiple RP/multiple PRD type of PMN imputation.

In the multiple RP/multiple PRD type, multiple univariate prediction models are used. The standard approach to multivariate modeling, with a given set of outcome variables (including both discrete and continuous), is likely to be computationally intensive due to the volume of model parameters and the difficulty in specifying a suitable covariance structure. Following Little and Rubin's (1987) proposal of a joint model for discrete and continuous variables, and its implementation by Schafer (1997), it is possible to fit a pure multivariate model for multivariate imputation, but it would require making distributional assumptions. Moreover, because of the obvious problem of specifying an accurate probability distribution underlying survey data, none of the existing solutions take the survey design into account. In the multiple RP/multiple PRD type, a multivariate model is fitted by a series of univariate parametric models (including the polytomous case), such that variables modeled earlier in the sequence have a chance to be included in the covariate set for subsequent models in the sequence.

---

[30] There is one exception to the rule that provisional hot-deck steps involve univariate matching and univariate assignment. The provisional hot-deck step for cocaine and crack lifetime use utilizes multivariate assignment, since both variables are used in the subsequent PRD model for heroin. The delta constraint refers to both predicted means, but in the calculation of Mahalanobis distance, only the cocaine predicted mean is used. Therefore, with respect to matching, this is not strictly univariate or multivariate; it is a little of both. See Section 6.3.1.5.

[31] There are exceptions. In a few imputation sets that use PMN Type 2 or PMN Type 3 (single RP/multiple PRD), provisional hot-deck steps are not completed because the variables earlier in the sequence are not used as covariates for variables later in the sequence. This occurs for some of the imputation sets for health insurance (Chapter 10) and roster pairs (Chapter 11).

Figure 3.3 PMN Type 2: Multiple Response Propensity/Multiple Prediction

For variables imputed by PMN Type 2 and PMN Type 3 (single RP/multiple PRD), the order in which variables were modeled is of some importance because variables early in the sequence have the potential to be part of the set of covariates for variables later in the sequence, but variables late in the sequence cannot be used for modeling for the earlier variables because of missing values. Note that usually not all variables in the sequence were missing for a particular incomplete record. Nevertheless, models were developed for all the variables in a univariate fashion for reasons mentioned earlier. For the drugs, the sequence of imputation was determined by considering such factors as the level of stigma associated with the drugs, the level of "missingness" in the data (Appendix A), and the degree to which one set of drugs could be used as predictors for other drugs. The decisions on sequencing for other imputation sets were made using similar criteria. For some respondents, some but not all of the variables in the imputation set are missing. This gives rise to missingness patterns (Section 3.3.3.4). Typically, in the final hot-deck step, only the predictive mean vector elements corresponding to missing variables are used to match donors with item nonrespondents. However, likeness constraints (and sometimes logical constraints) are often used to preserve relationships between the missing and nonmissing variables. Although the nonmissing values would not be replaced by the corresponding values from the donor, some degree of correlation between missing and nonmissing variables is expected to be preserved using these constraints.

The multiple RP/multiple PRD type works well for closely related variables that have different domains and different nonresponse patterns, because the separate RP steps account for these. The recency and frequency variables provide a good example: the domain of the recency models consists of all lifetime users; the domain of the 12-month frequency model (if applicable)

consists of all past year users; and the domain of the 30-day frequency model (if applicable) consists of all past month users. The provisional imputation-revised values may be used as covariates in later models, or even may be used to define the domains of later RP models.

3.4.3 Type 3: Single Response Propensity/Multiple Prediction

In PMN Type 3, the single RP/multiple PRD type, a single RP model is applied to all the variables modeled in the PRD steps. This is a less refined version of the preceding type, because it involves the fitting of only one RP model and is not as sensitive to different domains and response patterns among the outcome variables. The same weights are used for all PRD models. The lifetime drug use variables and source-of-income variables are examples of imputation sets that follow this type. Figure 3.4 illustrates the single RP/multiple PRD type of PMN imputation.

Figure 3.4 PMN Type 3: Single Response Propensity/Multiple Prediction

3.5 Special Auxiliary Variables: Age Group and State Rank

The age group and State of residency auxiliary variables apply to several of the imputation sets described in Chapters 4 through 11. Across variable groups, most imputation sets are processed separately by age group, regardless of the type of PMN that was used. The State of residence is used to construct a State-rank variable, which is then used in imputation for the drug variables (Chapter 6) and the income variables (Chapter 9).

3.5.1 Age Groups

The variables related to drug use, household composition, income, and health insurance are highly correlated with age. This, along with the desire to use parallel processing to expedite the time it takes to impute all the variables, led to the decision to separate the imputation procedures for these variables into distinct age groups. Therefore, the drug use variables were imputed within each of three age groups: 12 to 17, 18 to 25, and 26 or older. The household composition (roster-derived), income, and health insurance variables were imputed within the following four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older.[32] The roster pair variables (i.e., the variables related to the relationship between two respondents from the same sampled dwelling unit) were often divided into age groups depending on the ages of both pair members.

In the hot-deck step, the age group restriction could be considered a likeness constraint. However, the models also were built separately within the age groups, so this restriction was not loosened unless no other options were available. Although the demographic variables did not always show a high correlation with age, the imputation of missing values in the demographic variables also was performed within age groups. This was done to maintain consistency with how the other variables were imputed and to facilitate parallel processing. The same three age groups that were used for drugs were also used for demographics. Occasionally, small sample sizes necessitated the aggregation of age groups at the modeling stage. In particular, the models for education level (highest grade completed) were fit within the age groups of 12 to 17 and 18 or older. In the employment status models, the 15-17 and 18-25 age groups were aggregated. Finally, all age groups were aggregated for the Hispanic/Latino group, marital status, and immigrant age-of-entry models.

3.5.2 State Rank

Because State-level estimates are an important product of the NSDUH, there has been interest in requiring the donor to be from the same State as the recipient. However, this could not always be implemented because of insufficient pools of donors.[33] In such cases, information about the State of residence of each respondent was incorporated into the modeling and hot-deck steps of the PMN procedure by grouping respondents into three categories based on the ranking of their State of residence. For lifetime drug use, the States were ranked by the weighted proportion of lifetime users of the drug of interest. For recency and frequency of drug use, the States were ranked by the weighted proportion of past month users of the drug of interest. For income, the States were ranked by the weighted proportion of respondents whose personal incomes during the prior calendar year were greater than or equal to $20,000. These State-rank variables were used as covariates in the RP and PRD steps and sometimes in likeness constraints in the hot-deck step.

---

[32] Age groups were sometimes aggregated for the health insurance procedures. See Chapter 10 for details.

[33] In the hot-deck step for some of the demographic variables, a likeness constraint required the donor to be from the same segment as the item nonrespondent. Segments never cross State lines, so this can be viewed as a refined use of the State of residence. In practice, this constraint often had to be removed because many segments included only a handful of unit respondents.

4. Editing and Imputation for the NSDUH Core Demographics Variables

4.1 Introduction

As noted in Chapter 1, the NSDUH questionnaire includes both "core" and "noncore" modules. Questions about demographics are included in both of these types of modules. Core demographic questions include age, birth date, gender, marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level (highest grade completed). These questions remain in the survey each year and are important for trend estimates.

This chapter discusses procedures for editing and imputing the core variables for demographic characteristics. Although the interview date was not classified as a core demographic variable, it is used along with a respondent's birth date (when provided by the respondent) as the starting point for determining the respondent's current age. Therefore, the editing procedures for the interview date also are included in this chapter. The noncore demographic variables (including but not limited to immigrant status and employment status) are discussed in Chapter 5.

Prior to imputation, editing was performed on all of these core demographics variables. This editing could range from simply assigning legitimate skip codes (see Section 2.4.2 in Chapter 2), as was the case for marital status, to coding "OTHER, Specify" responses (see Section 2.3.1.3) and resolving inconsistencies, as was the case for race and ethnicity.

After editing, the variables were processed in one of four ways:

- No imputation required: interview date, age, gender. These are described in Sections 4.2.1, 4.2.2, and 4.2.4, respectively. The edited data contained no missing values.
- No imputation performed: current State residency reported by interviewers,[34] number of times married, military status, perceived health status.[35] These are described in Sections 4.2.3, 4.2.6, 4.2.8, and 4.2.10, respectively.
- Random assignment: birth date. This procedure is summarized in Section 4.2.5 because it is straightforward and does not involve the predictive mean neighborhood (PMN) method, which is described in Chapter 3.
- PMN: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, education level. These are described in Sections 4.2.6, 4.2.7, and 4.2.9, respectively.

[34] The State residency that was reported by interviewers differs from the variable STATE, which contains the Federal information processing standards (FIPS) codes for States and is created at the sampling stage. The variable STATE is recommended for use in State-level analyses.

[35] A final imputation-revised variable is not created for military service. However, a limited non-PMN imputation is performed for military service as part of the imputation procedures for health insurance, as described in Section 10.3.2.4.

Overall, the core demographics variables that are discussed in this chapter either did not have missing data by design or tended to have high item response rates. Except for race, the item response rates when nonresponse was allowed (i.e., "don't know" or "refused") were more than 99 percent.[36]

4.2 Editing Selected Core Demographics Variables

In this section, editing procedures are described for the following:

- interview date (Section 4.2.1);
- age (Section 4.2.2);
- State residency (Section 4.2.3);
- gender (Section 4.2.4);
- birth date (Section 4.2.5);
- marital status and number of times married (if respondents had ever been married) (Section 4.2.6);
- race, Hispanic/Latino origin, and Hispanic/Latino group (Section 4.2.7);
- U.S. military service history and current military status (if respondents had ever been in the U.S. military) (Section 4.2.8);
- education level (Section 4.2.9); and
- perceived health status (Section 4.2.10).

4.2.1 Creating the Edited Interview Date Variable (INTDATE)

The program for the computer-assisted interviewing (CAI) captured information about the time and date at the start of the interview, at the start of each module to which respondents were routed, and at the completion of the interview. This information was recorded in separate variables that were included in the interview data that were transmitted from the field to the data processing staff. These variables containing the start and end dates and times of the interview and modules are referred to as time and date "stamps." For example, the time and date stamps for the beginning of the interview were captured in the variable TBEGINTR, the time and date stamps for the end of the interview were captured in the variable TENDINR, and the time and date stamps for the alcohol questionnaire module were captured in the variable TBEGALC. These time and date stamps were sequentially arranged in the dataset in the order that the questionnaire modules appeared. In most cases, the time and date stamps were compared and edited to obtain the final edited interview date (INTDATE). The editing indicator associated with the interview date variable (EIIDATE) specified the module date stamp that was used to create the edited interview date.

[36] When given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group such as "Mexican," resulting in a comparatively larger amount of missing data for the race question. As a result, the item response rate for the race variables is still high but tends to be about 96 to 97 percent.

In some situations a "breakoff" occurred (i.e., the respondent did not finish the interview in one sitting and resumed the interview on another date), and the respondent's birthday occurred between the beginning and the end of the interview. In these situations, the interview date was set to the end-of-interview date stamp, which was consistent with the first date stamp after the respondent's birthday. (This date stamp was indicated in the CAI.)

A date stamp was not used to set the interview date if any of the following conditions were true:

1. The date stamp was more than 14 days outside the quarter in which the interview was supposed to take place.
2. The date stamp was later in time than a subsequent date stamp.
3. The date stamp occurred before a birthday, which in turn occurred before the end of the interview.

If none of the date stamps were usable, then the interview date was imputed to be sometime in the quarter in which the interview was assumed to have taken place.

4.2.2 Creating the Edited Age Variable (AGE)

The age that respondents reported at the beginning of the interview (CALCAGE) was determined in one of two ways:

1. The respondent reported his or her birth date to the interviewer. This information was captured in the variable AGE1. The interviewer confirmed with the respondent that the birth date information had been recorded correctly and could not proceed further with the interview until the respondent verified the birth date. CALCAGE was determined from AGE1 and the date that was recorded at the start of the interview. The interviewer then confirmed with the respondent that this calculated age was correct. If the respondent indicated that the calculated age was not correct, then the interviewer could not proceed further until any additional corrections to the birth date had been made and the respondent verified both the date of birth as well as the calculated age.
2. If the respondent did not know or refused to report a birth date, the respondent was asked to provide his or her correct age, which was captured in the variable DKREFAGE. In this situation, CALCAGE was set to the age that had been recorded in DKREFAGE.

Interviews were terminated if respondents did not report their age in one of these ways. Interviews also were terminated if CALCAGE indicated that respondents were less than 12 years old.

Even after CALCAGE has been determined for respondents at the beginning of the interview, respondents could change their age in response to consistency checks in self-administered modules pertaining to their substance use (e.g., if they reported first using a drug at an age that was greater than their current age). Any changes that respondents made to their age

during the questionnaire were captured in the temporary variable CURNTAGE. Therefore, it was possible for the age that was recorded for the respondent at the beginning of the interview (CALCAGE) to be different from the value of CURNTAGE that was captured at the end of the interview and stored in the variable NEWAGE. As discussed before, NEWAGE is the final value stored in the variable CURNTAGE at the end of the interview.

The final age variable, AGE, was determined using CALCAGE and NEWAGE and three other sources: the age calculated from the final edited interview date (INTDATE) and the reported birth date (AGE1), the age corresponding to the "self" in the questionnaire household roster (if it existed) (see Section 8.2.3), and the pre-interview age that was reported during screening of the dwelling unit (DU).[37] In most situations when the final edited continuous age was determined, priority was given to CALCAGE, NEWAGE, and the age that was calculated from AGE1 and INTDATE. There were occasions, however, where the age corresponding to the "self" in the household roster was used, even if it did not agree with CALCAGE and NEWAGE. If the final age (AGE) did not agree with the date of birth that had originally been entered (AGE1), the birth date also was edited. An intermediate value for age was determined in the following manner:

Intermediate value for age =

- NEWAGE, if nonmissing and exactly equal to CALCAGE, where TBEG_TUT (the interview date time stamp at the beginning of the tutorial) = INTDATE (the edited interview date) (age indicator = 1); else
- NEWAGE, if nonmissing, TBEG_TUT and INTDATE were not equal, but NEWAGE was exactly equal to CALCAGE (adjusted by Blaise[38] to a changed interview date if the interview date was changed within the questionnaire), and the respondent's birthday did not fall between the dates corresponding to TBEG_TUT and INTDATE (age indicator = 1); else
- NEWAGE, if nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, the given value of CALCAGE agreed with what it should be based on INTDATE and the given birth date (i.e., EIIDATE not equal to 6), and NEWAGE and CALCAGE were exactly equal (age indicator = 1); else
- age calculated from INTDATE and the reported birth date, if the birth date was nonmissing, TBEG_TUT and INTDATE were not equal, the respondent's birthday fell between the dates corresponding to TBEG_TUT and INTDATE, and the given value of CALCAGE did not agree with what it should be based on INTDATE and the given birth date (EIIDATE = 6), where the newly calculated age based on INTDATE

[37] When contacting the DU, the field interviewer (FI) asked to speak with an adult resident of the household aged 18 or older who could serve as the screening respondent. Using a handheld computer, the FI completed a 5-minute procedure with the screening respondent that involved listing all household members along with their basic demographic data (including age). The computer used the demographic data in a preprogrammed selection algorithm to select zero to two sample persons, depending on the composition of the household.

[38] Blaise is the computer program within the CAI instrument that was used to direct the respondent and interviewer through the questionnaire.

was exactly equal to the screener age and/or the roster age (if it existed) (age indicator = 2); else

- NEWAGE, if NEWAGE differed from CALCAGE and NEWAGE = screener age and NEWAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 3); else
- CALCAGE, if CALCAGE differed from NEWAGE and CALCAGE = screener age and CALCAGE = roster age (if it existed), and the interview date at the beginning of the interview (TBEGINTR) was within the appropriate quarter (age indicator = 4); else
- age calculated from reported birth date and INTDATE, if EIIDATE = 5 and NEWAGE = CALCAGE (but neither was equal to the correct age) (age indicator = 5); else
- NEWAGE, if NEWAGE differed from CALCAGE, but NEWAGE = roster age, provided roster age existed (age indicator = 6); else
- CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = roster age, provided roster age existed (age indicator = 7); else
- NEWAGE, if NEWAGE differed from age calculated from reported birth date and INTDATE, but NEWAGE = CALCAGE, screener age, and roster age (if it existed) (age indicator = 8); else
- CALCAGE, if CALCAGE differed from NEWAGE, but CALCAGE = age calculated from INTDATE and the reported birth date, and CALCAGE was within 1 year of screener age and roster age (age indicator = 9).

After these rules had been applied, this intermediate age value was compared with the age corresponding to the "self" in the household roster. For most respondents, the final edited value for the age variable (AGE) was set to this intermediate age value. There were exceptions, however, as detailed below.

By the time the interviewer reached the roster part of the questionnaire, there had been multiple opportunities to change the value of age in response to consistency checks. This value of age was called CURNTAGE by the Blaise program. One of the consistency checks in the questionnaire household roster was to verify the value of the respondent's own entry for age in the household roster (the "self" entry) against the value of CURNTAGE. If the self age differed from CURNTAGE, then the interviewer could either change the respondent's age that was entered in the roster or override the consistency check and provide an explanation for why the roster age did not match CURNTAGE. If the consistency check for age was overridden, then the value for age corresponding to the "self" in the roster may not match the intermediate age value that was described previously. However, if sufficient explanation was given for overriding the age consistency check, other evidence pointed to the veracity of the roster age, and the difference between CURNTAGE and the roster age for self was less than 2 years, then AGE was set to the roster age, even if it disagreed with both NEWAGE and CALCAGE. In particular, all of the following conditions had to be met for this to occur:

1. The interviewer specifically indicated that the roster age was the correct one.
2. The pre-interview screener age matched the roster age.
3. The other household member's roster supported the roster age value, if another member of the household completed the interview.

It also was possible for interviewers to jump back from the household roster or elsewhere in interviewer-administered sections (i.e., after respondents had completed the audio computer-assisted self-interviewing [ACASI] section of the interview) to change the respondent's date of birth. In this situation, the values of CALCAGE and NEWAGE could be consistent, but these changes to the respondent's age between the beginning and end of the interview would not be readily apparent. Rather, this pattern would become evident in reviewing ACASI data. For example, if respondents were aged 21 or older at the beginning of the interview and they did not change their age during the ACASI portion of the interview, questions were skipped in the noncore consumption of alcohol module about alcohol use by persons aged 12 to 20. However, if their final age indicated they were aged 12 to 20, then they would have been eligible to be asked questions about alcohol use by persons aged 12 to 20 (see Sections 7.3 and 7.4.17 in Chapter 7).

Three age category variables were created from the final age: CATAGE with four levels (12 to 17, 18 to 25, 26 to 34, and 35 or older); CATAG2 with three levels (12 to 17, 18 to 25, and 26 or older); and CATAG3 with five levels (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older). These variables were used instead of the continuous age variables in some subsequent imputations and analyses.

4.2.3 Creating the Edited Current State Residency Variable (STATELOC)

Field interviewers (FIs) were instructed to use the FI "checkpoint" FIPE4 at the beginning of the interview to report the State where the DU was located. The term "checkpoint" refers to an item that was completed by the FI about the location of the DU or characteristics of the sample within the DU. These checkpoints were not administered to the respondents but were used to customize the wording of questions in subsequent sections of the interview, such as State-specific names for Temporary Assistance for Needy Families (TANF) programs (see Section 9.2.3).

The State that interviewers entered in FIPE4 sometimes failed to match the State residence information that was used to sample a given case. These mismatches were investigated by field staff during data collection. Some of these mismatches existed for a valid reason, such as if a respondent had been selected in a DU in one State but had moved to another State. In these situations, if FIPE4 reflected the State where the respondent was currently living, the edited variable STATELOC (corresponding to FIPE4) retained the value from FIPE4. Otherwise, if the State information in FIPE4 was entered incorrectly, STATELOC was assigned a code for bad data (see Section 2.4.3). In turn, the setting of STATELOC to bad data affected the editing of other variables that were dependent on the State where the respondent was reported to be a resident.

4.2.4 Creating the Edited Gender Variable (IRSEX)

As with previous surveys since 2002, FIs in 2012 were required to enter the respondent's gender in question QD01. As a result, it was not possible to have missing values for this question. To maintain continuity with surveys prior to 2002 in which missing data could exist for gender, the variable name IRSEX was used to describe gender in the 2012 survey. However, it was not necessary to create an imputation indicator, because IRSEX and QD01 were equivalent.

As for the situation described previously for age, interviewers also could jump back to the beginning of the interview and change the respondent's gender based on roster information. This type of change to QD01 between the beginning and end of the interview would not be readily apparent. However, this pattern would become evident in reviewing ACASI data. For example, if respondents were defined as being male at the beginning of the interview, questions in the noncore health care module about pregnancy were skipped. However, if their final gender and age indicated that they were female and aged 12 to 44, then they would have been eligible to be asked questions about pregnancy (see Sections 7.3 and 7.4.9 in Chapter 7).

4.2.5 Creating the Edited Birth Date Variable (BRTHDATE)

To proceed with the interview, respondents were required to provide their date of birth or current age (if they did not provide their date of birth) at the beginning of the interview. Each completed case respondent possessed a current age, although a number of cases had missing birth dates. If the birth date was nonmissing but was inconsistent with AGE and INTDATE (either in the unedited data or as a result of editing age and/or interview date), then the reported birth month and day were preserved, and the birth year was logically edited according to the interview date and age.

In cases with missing birth dates, a birth date was randomly selected from all possible birth dates, given the final age and interview date. Each date in this period (365 or 366 days, depending on whether the period includes February 29 in a leap year) had an equal probability of selection.

4.2.6 Creating the Edited Marital Status Variables (MARITAL, EDMARIT, and NOMARR)

In the 2012 questionnaire, a single core question (QD07) asked about the respondent's marital status if respondents were aged 15 or older. If respondents indicated that they were currently married, widowed, or divorced or separated, they also were asked to report the number of times they had been married (QD08). The exact phrasing of the questions was as follows:

QD07: Are you now married, widowed, divorced or separated, or have you never married?

1 MARRIED
2 WIDOWED
3 DIVORCED OR SEPARATED
4 HAVE NEVER MARRIED

QD08: [IF QDO7 = 1 OR 2 OR 3] How many times have you been married?

NUMBER OF TIMES: _____ [RANGE: 1 - 9]

MARITAL and NOMARR underwent minimal processing. Legitimate skip codes were assigned to both variables (Section 2.4.2) if respondents were aged 12 to 14. Similarly, legitimate skip codes were assigned to NOMARR if respondents had never been married. The base variable for creating an imputation-revised version of marital status was called EDMARIT. This variable was equivalent to MARITAL, with the exception that all legitimate skip codes were collapsed into a single legitimate skip code (99), and missing values were set to the SAS[39] missing code (.) so that they could be properly handled by the modeling programs.

4.2.7 Creating the Edited Race and Hispanic/Latino (Origin and Group) Variables

In the 2012 questionnaire, two core questions focused on the respondent's ethnicity[40] (QD03 and QD04) and two focused on the respondent's race (QD05 and QD05ASIA). For those questions with multiple categories (QD04, QD05, and QD05ASIA), the respondent had the opportunity to select more than one category. Two more Hispanic/Latino group categories were added to QD04 since the 2004 survey: Dominican (from Dominican Republic) and Spanish (from Spain). These new categories were added to the survey because of the large number of "OTHER, Specify" responses in previous NSDUHs that mapped to these categories.

The questions as they appear in the survey instrument are presented below.

QD03: Are you of Hispanic, Latino, or Spanish origin or descent?

1 YES
2 NO

QD04: (Asked only if QD03 = 1) Which of these Hispanic, Latino, or Spanish groups best describes you?

1 MEXICAN / MEXICAN AMERICAN / MEXICANO / CHICANO
2 PUERTO RICAN
3 CENTRAL OR SOUTH AMERICAN
4 CUBAN / CUBAN AMERICAN
5 DOMINICAN (FROM DOMINICAN REPUBLIC)
6 SPANISH (FROM SPAIN)
7 OTHER (SPECIFY)

QD05: Which of these groups describes you?

1 WHITE
2 BLACK / AFRICAN AMERICAN

[39] SAS® software is a registered trademark of SAS Institute Inc.

[40] The questions about ethnicity were limited to determining whether a respondent was Hispanic/Latino or not, and the specific Hispanic/Latino group to which a Hispanic/Latino respondent belonged.

3 AMERICAN INDIAN OR ALASKA NATIVE (AMERICAN INDIAN INCLUDES NORTH AMERICAN, CENTRAL AMERICAN, AND SOUTH AMERICAN INDIANS)
4 NATIVE HAWAIIAN
5 OTHER PACIFIC ISLANDER
6 ASIAN (FOR EXAMPLE: ASIAN INDIAN, CHINESE, FILIPINO, JAPANESE, KOREAN, AND VIETNAMESE)
7 OTHER (SPECIFY)

QD05ASIA: (Asked only if level 6 of QD05 was selected) Which of these groups describes you?

1 ASIAN INDIAN
2 CHINESE
3 FILIPINO
4 JAPANESE
5 KOREAN
6 VIETNAMESE
7 OTHER (SPECIFY)

As stated in the guidelines from the Office of Management and Budget (OMB),[41] "Hispanic/Latino" was categorized as an ethnicity, not a race. However, when given the opportunity to enter a race, many respondents entered "Hispanic" or some Hispanic/Latino group, resulting in missing data for the race question. Even though the final drug use tables were cross-classified with a variable that combined race and ethnicity, separate variables were initially created for race and ethnicity, and the race/ethnicity variables used in the tables were derived from these separate variables.

Due to the relationship between Hispanicity and race reporting, Hispanicity was used in the editing of race, and vice versa. In the process of editing race, the "OTHER, Specify" response to the Hispanic/Latino group question (QD04) was consulted (if it existed) if no race information was identified in QD05 or QD05ASIA. Similarly, in the process of editing the Hispanic/Latino group, the "OTHER, Specify" responses to the race questions (QD05 and QD05ASIA) were consulted (if they existed) if no Hispanic/Latino group information was identified in QD04. Because of the interdependence of race and Hispanicity, the editing of these variables is discussed together in this section.

The procedures used to edit the race and Hispanicity variables in the surveys since 2008 differed in several ways from the procedures used in previous surveys. One of the major differences was in the handling of race for multiple-race respondents. The first procedural changes were triggered by the elimination of the QD06 question, which appeared in the survey from 1999 to 2002. QD06 asked respondents who selected more than one racial category from QD05 and QD05ASIA combined to choose the race with which they identified the most. Without this question, it was impossible to determine (directly) the single race that a given

[41] In October 1997, the OMB released a notice, "Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity" (OMB, 1997) that provides new standards for maintaining, collecting, and presenting Federal data on race and ethnicity.

multiple-race respondent would most closely identify for himself or herself. In the 2003-2007 surveys, QD06 responses were "simulated" based on models built using true QD06 responses from the 2000-2002 surveys.[42] However, because racial demographics in the United States had changed since the 2000 survey and because recent data that were needed to update these models were not available, this method was not used after 2008 and single races were not assigned for multiple-race respondents. Refer to Section 3.3 of the 2008 imputation report (Ault et al., 2010) for more details.

4.2.7.1 Categories Used in Race and Hispanic/Latino Variables

4.2.7.1.1 Race Categories

For editing purposes, the 5 specific categories in QD05 (white, black/African American, American Indian/Alaska Native, Native Hawaiian, and Other Pacific Islander) and the 6 specific categories in QD05ASIA (Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese) were combined to produce 11 racial categories. Two other categories also were created: "Other Asian" (where the responses to QD5ASIA did not fit into the above category) and "Asian nonspecific" (where no response was selected to QD05ASIA, even though Asian was selected in QD05). Respondents could choose almost any subset of these categories. The only subsets that were not logically possible were those that included "Asian nonspecific" in combination with one or more specific Asian categories. Combining the information from QD05 and QD05ASIA, as well as QD04 when necessary, allowed the creation of all the edited and imputation-revised race variables.

4.2.7.1.2 Hispanic/Latino Categories

With the addition of two Hispanic/Latino categories since the 2004 survey, respondents were given the choice of seven categories in QD04 (Mexican/Mexican American/Mexicano/Chicano, Puerto Rican, Central or South American, Cuban/Cuban American, Dominican (from Dominican Republic), Spanish (from Spain), or some other Hispanic/Latino group),[43] and they could choose more than one category. As with QD05, interviewers could manually enter the alternative to the choices given, which would be either coded to some subset of the existing seven categories or set to missing. The "OTHER, Specify" responses to QD05 and/or QD05ASIA, if nonmissing, were consulted if no Hispanic/Latino origin group information was available from QD04. The final imputation-revised Hispanic/Latino group variable, IRHOGRP4, included all seven Hispanic/Latino group levels and a legitimate skip code (99) for respondents who were not Hispanic/Latino.

---

[42] Because of the questionnaire differences between the 1999 survey and the 2000-2002 surveys, the procedure for simulating QD06 responses for the 2003-2007 surveys was made simpler by limiting the QD06 results from the 2000-2002 surveys. During the 2003-2007 surveys, for the purpose of allocating respondents into imputation classes, a model was used to select a single race for respondents who had selected more than one race (IRRACE2). The selection of a single race was based on models that were fit using data from the 2000-2002 surveys. This method is described in Appendix E of the 2007 imputation report (Ault et al., 2009).

[43] When listing the six Hispanic/Latino defined categories in QD04, they shall henceforth be listed in this chapter as Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spanish.

4.2.7.2 Classification of "OTHER, Specify" Codes

All "OTHER, Specify" responses from QD04, QD05, and QD05ASIA were assigned both a race code and a Hispanic/Latino code. Each of these codes was mapped to at least one of the categories described in Section 4.2.7.1 and in this section, or to some other code that was informative in the final imputation described in Sections 4.3.2 and 4.3.3. A summary of categories of "OTHER, Specify" codes and how they were handled is given in the following sections. Appendix C provides the individual "OTHER, Specify" codes and more details about how they were handled.

4.2.7.2.1 Mapping of Race "OTHER, Specify" Codes to Edited Values

This section describes the directly and indirectly mapped race codes. The edits following from either of these types of mapped codes resulted in values that were considered "final" in that no imputation was necessary for them.

The directly mapped codes were mapped to one or more of the categories given in the questionnaire (see Section 4.2.7). There were directly mapped racial category codes and directly mapped geographic category codes. Racial category codes were exactly equivalent to one or more categories in QD05 or QD05ASIA, and were mapped directly to those categories regardless of whether the write-in response was in QD05 or QD05ASIA. (Respondents were still considered at least part Asian, even if the write-in response in QD05ASIA was non-Asian. The racial makeup of a respondent who entered a non-Asian racial category in QD05ASIA was determined on a case-by-case basis.) For example, a response such as "Han" mapped directly to a category in QD05ASIA ("Chinese"), and a response such as "mestizo" mapped directly to two categories in QD05, "white" and "Native American."

By contrast, geographic category codes corresponding to a country where census data indicated a racially homogeneous society depended on the corresponding question. For example, an entry of "Polish" in QD05 mapped to white because the Polish census data indicated nearly all Poles were white. On the other hand, an entry of "Polish" in the QD05ASIA "OTHER, Specify" mapped to "Other Asian." Geographic category codes also included ethnic groups where the racial identification was not immediately obvious. For example, a response of "Arab" would be automatically mapped to "white" if the response was a write-in response for QD05. However, as with the "Polish" entry, if the "Arab" response was a write-in response in QD05ASIA, the respondent was considered "Other Asian."

Indirect mapping was used for countries that were racially heterogeneous. A racial category was chosen by generating a random number and allocating the race based on a comparison of the random number with the proportions of races in the country's census.[44] For example, an entry of "Bolivian" would have a 55 percent chance of being allocated to the American Indian/Alaska Native category, because the latest Bolivian census indicated 55 percent of Bolivians were American Indian/Alaska Native. For countries where the census indicated a small proportion of some indistinct category such as "other" and the randomly generated number

[44] See www.infoplease.com/ipa/A0855617.html (Ethnicity and Race by Countries) for more information.

indicated an allocation to this proportion, the final race was left to imputation (appropriately constrained based upon the indistinct response).

If two or three heterogeneous countries were entered in the "OTHER, Specify" response (e.g., Bolivian and Peruvian), the final race was allocated using the following procedure: (1) randomly assign races based on the proportions for each country mentioned; and (2) combine the results. Exceptions to these rules occurred with the categories Mexicans, Puerto Ricans, Cubans, Dominicans, Central or South Americans (no country listed), and Spanish, which were given codes described under the next heading, with a final value determined using the formal imputation procedures described in Section 4.3.2. Starting with the 2006 survey, the imputation processing of indirectly mapped codes obtained from QD05ASIA has been simplified. In prior survey years, this type of write-in response was mapped to a race through country census information; since the 2006 NSDUH, all census-based write-in responses to the Asian race question were mapped directly to the "Other Asian" racial category.

4.2.7.2.2 Mapping of Race "OTHER, Specify" Codes to Inform Imputation

"OTHER, Specify" responses that could not be mapped definitively to a specific race category resulted in incomplete values requiring imputation. These responses were assigned two types of codes, either informative or noninformative, for the formal imputation procedures for race described in Section 4.3.2.

Responses that provided information were used to limit the final imputation. For example, a response of "mixed" resulted in an imputation among donors with two or more races, and a response of "brown" resulted in an imputation among donors who were not single race white.

A noninformative response (e.g., American) that was not accompanied by a response to one of the precoded categories[45] in QD05 or QD05ASIA (i.e., those other than the "OTHER, Specify" response) resulted in an unrestricted imputation.

4.2.7.2.3 Subsequent Editing of Race "OTHER, Specify" Codes

Subsequent to the initial mapping of the race "OTHER, Specify" codes, edits were sometimes implemented that revised or clarified the initial mapping before final races were allocated. These edits were necessary if multiple sources of information, including "OTHER, Specify" responses, provided conflicting or confusing information. These edits were implemented when (1) the final mapping depended upon the source question (i.e., QD04, QD05, and QD05ASIA); (2) the responses were given to both the "OTHER, Specify" and precoded categories of QD05 or QD05ASIA; or (3) the different "OTHER, Specify" responses were present in at least two of QD04, QD05, and QD05ASIA. In some cases, it was necessary to individually examine the responses to determine the appropriate mapping.

[45] For example, precoded categories in QD05 in 2012 were 1 = White; 2 = Black/African American; 3 = American Indian or Alaska Native; 4 = Native Hawaiian; 5 = Other Pacific Islander; and 6 = Asian. Category 7 (Other) led to the "OTHER, Specify" question.

Occasionally, the final mapped value depended upon whether the "OTHER, Specify" code was in QD04, QD05, or QD05ASIA. An example from directly mapped codes is "Indian." This response would be mapped to "American Indian/Alaska Native" if the "OTHER, Specify" response was in QD05, but it would be mapped to "Asian Indian" if the "OTHER, Specify" response was in QD05ASIA. Indirectly mapped codes also could depend upon the source question. The census data from many countries included Asian categories. If the "OTHER, Specify" response was in QD05ASIA, the random imputation to a census category was limited to the Asian categories. "OTHER, Specify" responses that were not specifically Asian sometimes occurred in the "OTHER, Specify" category of QD05ASIA. These were carefully examined, but the "Asian" part of the response was always preserved.

If "OTHER, Specify" responses to QD05 or QD05ASIA accompanied responses to the precoded categories in QD05 and QD05ASIA, it was necessary to reconcile these responses. In some cases, the combination of responses mapped to one of the multiple racial categories. For example, if a respondent selected "black/African American" in QD05 and wrote in "black and American Indian," then the respondent would be assigned both racial categories "black/African American" and "American Indian/Alaska Native."

There were instances, however, when the "OTHER, Specify" response was ignored because of responses to the precoded categories in QD05 and QD05ASIA. In particular, the "OTHER, Specify" response was always ignored if a precoded response category was selected, and the "OTHER, Specify" response was a geographic category code.[46] For example, if the interviewer selected the category for "black/African American" for the respondent and also wrote in "Polish," it was assumed that the respondent was a black Pole and, for racial identification purposes, was considered single-race black/African American. This was true even though the Polish census did not identify significant numbers of nonwhite persons in the Polish population.

In some instances, it was necessary to reconcile the "OTHER, Specify" responses to QD04, QD05, and QD05ASIA. In these cases, the responses were examined on an individual basis, and sometimes a new code was assigned that more accurately reflected the situation.

4.2.7.2.4 Mapping of Hispanic/Latino "OTHER, Specify" Codes

Certain Hispanic/Latino codes were considered "Definitely Hispanic." If any of these appeared in QD05 or QD05ASIA, the respondent was considered Hispanic/Latino regardless of the response to QD03. Examples included "Hispanic" and "Dominicano" (Spanish for "Dominican"). There was also a code to handle respondents who were definitely not Hispanic/Latino (i.e., the respondent reported "Not Hispanic/Latino"). If this code appeared in QD04, QD05, or QD05ASIA, then the respondent was considered non-Hispanic/Latino regardless of the response to QD03. All other Hispanic/Latino codes either mapped directly to one or more of the seven Hispanic/Latino group categories or provided no new information (e.g., Hispanic).

[46] Actually, this "edit" was not "subsequent" to the initial mapping. Instead, the initial mapping was ignored under the circumstances described.

4.2.7.3 Edited Race Variables

4.2.7.3.1 Individual Race Categories (EDQD051-EDQD0513)

Edited variables were created that correspond to the 13 racial categories described in Section 4.2.7.1.1. These variables were called EDQD05*xx*, where *xx* represented a number between 1 and 13, corresponding to each of the 13 categories.

EDQD05*xx* =

- 1, if the level *xx* was selected by the respondent in QD05 or QD05ASIA; else
- 2, if the level *xx* was indicated by a directly mapped code in QD05 or QD05ASIA; else
- 3, if no EDQD05*xx* variables had values of 1 or 2, and the level *xx* was indicated by a directly mapped code in QD04 (Hispanic/Latino status); else
- 4, if (a) no EDQD05*xx* variables had values of 1, 2, or 3, and (b) the level *xx* was indicated by an indirectly mapped code in QD04, QD05, and/or QD05ASIA; else
- missing.

EDQD0513 (Asian nonspecific) was a little different from the others. In particular, there was no specific level of QD05 or QD05ASIA that corresponded to it. It was used mainly to preserve a response of "Asian" to QD05, even if the respondent selected nothing in QD05ASIA. The value of EDQD0513 was set to 1 if the respondent selected "Asian" in QD05 but mentioned nothing that mapped to a specific Asian category in QD05ASIA. It also could have values of 2, 3, or 4, depending on the "OTHER, Specify" codes.[47]

4.2.7.3.2 Broad Categories of Race (EDRACE)

The EDRACE is a 24-level variable that indicates which of four broad racial categories (white, black/African American, American Indian/Alaska Native, Asian/Other Pacific Islander) were identified in QD04, QD05, and QD05ASIA, and it also has levels to indicate how the imputation should be restricted based on the race of the donor. The first three broad racial categories corresponded to EDQD051, EDQD052, and EDQD053, respectively. "Asian/Other Pacific Islander" was considered to have been identified if any of EDQD054 through EDQD0513 was nonmissing. EDRACE was created using the following rules, under five possible scenarios:

Scenario 1: If only one broad racial category was identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 1 (white only), if EDQD051 was nonmissing; else
- 2 (black/African American only), if EDQD052 was nonmissing; else

[47] A value of 2 indicated that the respondent wrote "Asian" in the QD05 other-specify blank. A value of 3 indicated that the response was obtained from the other-specify part of the Hispanic/Latino group question (QD04). Finally, a value of 4 indicated that the respondent gave a country of origin as a response to QD05, and the census for that country had "Asian" as one of its categories.

- 3 (American Indian/Alaska Native only), if EDQD053 was nonmissing; else
- 4 (Asian/Other Pacific Islander only), if any of EDQD054 through EDQD0513 were nonmissing.

Scenario 2: If two broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 5 (white and black/African American only), if both EDQD051 and EDQD052 were nonmissing; else
- 6 (white and American Indian/Alaska Native only), if both EDQD051 and EDQD053 were nonmissing; else
- 7 (white and Asian/Other Pacific Islander only), if EDQD051 was nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing; else
- 8 (black/African American and American Indian/Alaska Native only), if both EDQD052 and EDQD053 were nonmissing; else
- 9 (black/African American and Asian/Other Pacific Islander only), if EDQD052 was nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing; else
- 10 (American Indian/Alaska Native and Asian/Other Pacific Islander only), if EDQD053 was nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing.

Scenario 3: If three broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 11 (white, black/African American, and American Indian/Alaska Native only), if all of EDQD051 through EDQD053 were nonmissing; else
- 12 (white, black/African American, and Asian/Other Pacific Islander only), if both EDQD051 and EDQD052 were nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing; else
- 13 (white, American Indian/Alaska Native, and Asian/Other Pacific Islander only), if both EDQD051 and EDQD053 were nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing; else
- 14 (black/African American, American Indian/Alaska Native, and Asian/Other Pacific Islander only), if both EDQD052 and EDQD053 were nonmissing and at least one of EDQD054 through EDQD0513 were nonmissing.

Scenario 4: If all four broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE = 15.

Scenario 5: If none of the broad racial categories were identified in QD04, QD05, and/or QD05ASIA, EDRACE =

- 16 (multiple race, no other information), if an "OTHER, Specify" answer such as "biracial" or "mixed" appeared in QD04, QD05, or QD05ASIA; else
- 17 (nonwhite, no other information), if an "OTHER, Specify" answer such as "brown," "tan," or similar answers in Spanish appeared in QD04, QD05, or QD05ASIA; else
- 18 (white, or both white and American Indian/Alaska Native), if the random assignment of a census data code resulted in imputation restricted to donors who were either white, or both white and American Indian/Alaska Native; else
- 19 (not American Indian/Alaska Native, in part or in full), if the random assignment of a census data code resulted in imputation restricted to donors who were not American Indian/Alaska Native, in part or in full; else
- 20 (non-Hispanic Mexican), if "Mexican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 21 (non-Hispanic Cuban), if "Cuban" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 22 (non-Hispanic Central or South American), if "Central or South American" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 23 (non-Hispanic Dominican), if "Dominican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 24 (non-Hispanic Spanish), if "Spanish" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- 25 (non-Hispanic Spanish), if "Puerto Rican" was mentioned in the QD05 and/or QD05ASIA "OTHER, Specify" responses, but QD03 = 2; else
- missing.

4.2.7.3.3 Finer Categories of Race (EDNWRACE)

EDNWRACE was a 15-level edited variable used as a base variable for the imputation-revised finer racial category variable IRNWRACE. It also had a 16th level to indicate when the imputation should be restricted to Asian-specific categories. It was created using the following rules, under three possible scenarios:

Scenario 1: If only one of EDQD051 through EDQD0513 was nonmissing,

EDNWRACE =

- 16 (Asian nonspecific only), if EDQD0513 was the nonmissing variable; else
- *xx* (one known racial category only), where EDQD05*xx* was the nonmissing variable out of EDQD051 through EDQD0512.

Scenario 2: If more than one of EDQD051 through EDQD0513 was nonmissing,

EDNWRACE =

- 13 (Native Hawaiian and Other Pacific Islander only), if both EDQD054 and EDQD055 were nonmissing, and all other EDQD05*xx* variables were missing; else
- 14 (Asian multiple category), if all of EDQD051 through EDQD055 were missing (i.e., at least two of the ordinary Asian categories were selected); else
- 15 (more than one race).

Scenario 3: If all of EDQD051 through EDQD0513 were missing,

EDNWRACE =

- 15 (more than one race), if EDRACE = 16; else
- missing.

4.2.7.4 Edited Hispanic/Latino Variables

4.2.7.4.1 Hispanic/Latino Indicator (EDHOIND)

An edited Hispanic/Latino indicator, EDHOIND, was created using responses to QD03 and, in rare cases, the "OTHER, Specify" responses to QD04, QD05, and/or QD05ASIA. This indicator variable was created as follows:

EDHOIND =

- 1 (Hispanic/Latino), if QD03 = 1 and no "OTHER, Specify" response stated that the respondent was definitely not Hispanic/Latino, or if the "OTHER, Specify" response to QD05 or QD05ASIA indicated that the respondent was definitely Hispanic/Latino; else
- 2 (not Hispanic/Latino), if QD03 = 2 and no "OTHER, Specify" response stated that the respondent was definitely Hispanic/Latino, or if the "OTHER, Specify" response to QD04, QD05, and/or QD05ASIA indicated that the respondent was definitely not Hispanic/Latino; else
- missing.

The race "OTHER, Specify" responses, which were considered "definitely Hispanic/Latino," and the single Hispanic/Latino "OTHER, Specify" response, which was considered "definitely not Hispanic/Latino," are listed in Appendix C.

4.2.7.4.2 Individual Hispanic/Latino Group Categories (EDQD041-EDQD047)

The edited variables EDQD041 through EDQD047 were created to match the seven Hispanic/Latino group categories described in Section 4.2.7.1.2: Mexican, Puerto Rican, Central or South American, Cuban, Dominican, Spanish, and Other Hispanic/Latino.

EDQD04*xx* =

- 1, if the level *xx* was selected by the respondent in QD04; else
- 2, if the "OTHER, Specify" response from QD04 mapped directly to level *xx*; else
- 3, if no EDQD04*xx* variables had values of 1 or 2, and the "OTHER, Specify" response from QD05 or QD05ASIA mapped directly to level *xx*; else
- missing.

4.2.7.4.3 Edited Hispanic/Latino Group (EDHOGRP)

The edited variable EDHOGRP was the base variable for creating an imputation-revised Hispanic/Latino group variable. It had seven levels to match the seven Hispanic/Latino group categories described in Section 4.2.7.1.2, plus several other more general Hispanic/Latino levels that could be used in a restricted imputation. Those respondents with EDHOIND = 2 were assigned EDHOGRP = 99. It was created using the following rules, under four possible scenarios:

Scenario 1: If EDHOIND = 2,

EDHOGRP = 99.

Scenario 2: If EDHOIND = 1 or missing and only one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP = *xx*, where EDQD04*xx* was the nonmissing one.

Scenario 3: If EDHOIND = 1 or missing and more than one of EDQD041 through EDQD047 was nonmissing,

EDHOGRP =

- 1 (Mexican), if EDQD041 was nonmissing; else
- 2 (Puerto Rican), if EDQD042 was nonmissing; else
- 3 (Central or South American), if EDQD043 was nonmissing; else
- 4 (Cuban), if EDQD044 was nonmissing; else
- 5 (Dominican), if EDQD045 was nonmissing; else
- 6 (Spanish), if EDQD046 was nonmissing; else
- 7 (Other), if EDQD047 was nonmissing.

For the multiple Hispanic/Latino group respondents, a priority rule similar to the one used in the surveys prior to 2004 was applied in determining a single Hispanic/Latino group. The only difference is the addition of two more Hispanic/Latino group categories since the 2004 survey, resulting in the following order: Mexican, Cuban, Puerto Rican, Central or South American, Dominican, Spanish, and Other Hispanic/Latino.

Scenario 4: If EDHOIND = 1 or missing and all of EDQD041 through EDQD047 were missing,

EDHOGRP =

- EDRACE + 7 (imputation restricted by race), if 1 ≤ EDRACE ≤ 14; else
- missing.

4.2.8 Creating the Edited Military Service History and Current Service Variables (SERVICE and MILSTAT)

SERVICE and MILSTAT were created from QD09 and QD10, respectively, which asked respondents aged 17 or older about their current and past service in the U.S. military. The questions as they appear in the survey instrument are presented below.

QD09: Have you ever been in the United States' armed forces?

1 Yes
2 No

QD10: Are you **currently** on **active** duty in the armed forces, in a reserves component, or now separated or retired from either reserves or active duty?

1 On active duty in the armed forces
2 In a reserves component
3 Now separated or retired from either reserves or active duty

Respondents who were currently on active duty in the U.S. military were not eligible to be interviewed for NSDUH. If respondents reported in QD10 that they were currently on active military duty, the interviewers were asked to confirm this answer with the respondents. The interview was terminated if respondents confirmed that they were on active duty in the U.S. military. Consequently, there were no final respondents in the final NSDUH data who reported that they currently were on active military duty. However, some final respondents could be civilians who were currently in the military reserves or were separated or retired from the military. In addition, the industry and occupation variables in the noncore employment section could include military-related codes for some respondents (see Section 5.2.4)

Unlike the situation in most places in the interview (see Section 2.4.2), responses of "don't know" or "refused" to the question about lifetime military service were treated as potential indications of military service. Thus, respondents who did not know or refused to report whether they had ever been in the U.S. military also were asked QD10 in order to determine their eligibility status for the interview.

Legitimate skip codes were assigned to the SERVICE and MILSTAT if respondents were aged 12 to 16. In addition, legitimate skip codes were assigned to MILSTAT if respondents were aged 17 or older but they reported in QD09 that they had never been in the U.S. armed forces.

4.2.9 Creating the Edited Highest Grade Completed Variables (EDUC and EDEDUC)

EDUC and EDEDUC were created using the responses to the core education question QD11, which asked about the highest grade in school that the respondent had completed. The question from the survey instrument appears below:

QD11: What is the highest grade or year of school you have **completed**?

0	NEVER ATTENDED SCHOOL
1	1ST GRADE COMPLETED
2	2ND GRADE COMPLETED
3	3RD GRADE COMPLETED
4	4TH GRADE COMPLETED
5	5TH GRADE COMPLETED
6	6TH GRADE COMPLETED
7	7TH GRADE COMPLETED
8	8TH GRADE COMPLETED
9	9TH GRADE COMPLETED
10	10TH GRADE COMPLETED
11	11TH GRADE COMPLETED
12	12TH GRADE COMPLETED
13	COLLEGE OR UNIVERSITY / 1ST YEAR COMPLETED
14	COLLEGE OR UNIVERSITY / 2ND YEAR COMPLETED
15	COLLEGE OR UNIVERSITY / 3RD YEAR COMPLETED
16	COLLEGE OR UNIVERSITY / 4TH YEAR COMPLETED
17	COLLEGE OR UNIVERSITY / 5TH OR HIGHER YEAR COMPLETED

No editing was performed to create EDUC based on other questionnaire information. In particular, EDUC was not edited with respect to education variables in the noncore demographics section (e.g., current grade), nor was it edited with respect to the respondent's age. Consequently, the core education variable would not be affected by changes that might occur in the content of noncore education variables in subsequent years. A second variable, EDTEDUC, was created as part of the noncore demographics processing (see Section 5.2.3.3), but EDTEDUC was not used in any published estimates involving educational attainment.

The base variable for creating an imputation-revised version of education was called EDEDUC. It was equivalent to EDUC, except that missing values that described the type of item nonresponse (i.e., "don't know" or "refused") were set to the SAS missing code (.) so that they were properly handled by the imputation-related modeling programs.

4.2.10 Creating the Perceived Health Status Variable (HEALTH)

HEALTH was created from core question QD12, which asked the respondent to rate his or her health. The question from the survey instrument appears below:

QD12: This question is about your overall health. Would you say your health in general is excellent, very good, good, fair, or poor?

1 Excellent
2 Very Good
3 Good
4 Fair
5 Poor

No processing of this variable was done beyond that of assigning the edited variable name HEALTH instead of QD12 (the variable name obtained from the CAI).

4.3 Imputation for Selected Core Demographics Variables

In this section, the imputation procedures applied to the marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level are described. These variables comprised five independent manifestations of the single response propensity (RP)/single prediction (PRD) type of PMN described in Chapter 3. Each imputation set is discussed in a separate section below.

4.3.1 Marital Status Variable (Imputation Set 1)

The first core demographic variable that underwent imputation was the marital status variable. The four substantive levels of the imputation-revised variable IRMARIT matched the four answer categories of QD07 (i.e., married, widowed, divorced or separated, or never married). Respondents aged 12 to 14 were automatically assigned an IRMARIT value of 99, a "legitimate skip" code. Since this is the first variable to undergo imputation in each cycle, there were no imputation-revised variables to use as auxiliary variables. This tended to make the imputation process simple and straightforward.

In marital status imputation procedures, only one RP model and only one PRD model was fit; most other NSDUH imputation procedures are run separately within three or four age groups. Single models were used across all age groups to ensure adequate sample size for response categories that would be rare within certain age groups (e.g., the "widowed" category for younger age groups). To account for the correlation between age and marital status, AGE was used in both the RP and PRD model steps and in a likeness constraint in the hot deck step.

The marital status variable has a very high response rate (see Appendix A). There are often fewer than ten missing values in the entire sample in a given survey year.

4.3.1.1 Response Propensity Step

The response propensity model for imputation set 1 utilized the preliminary analysis weight, PANALWT. The marital status question QD07 was only asked of respondents aged 15 or older. Therefore, the domain contained unit respondents with AGE ≥ 15. The creation of the AGE variable is described in Section 4.2.2. Unit respondents in the domain with nonmissing EDMARIT values were considered item respondents. The EDMARIT variable is described in

Section 4.2.6. See Table D.1 in Appendix D for details of the covariates used in the RP model for this variable.

4.3.1.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the marital status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The outcome variable had four levels, which mapped to the four answer categories of QD07. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent selected each of the four answer categories of QD07.

4.3.1.3 Hot-Deck Step

The predicted means from the PRD step play a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps for marital status were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This was done to allow parallel processing, which decreases the time required for implementation. No logical constraints were used, and the only likeness constraint other than the delta constraint involved the continuous AGE variable. The few unit respondents requiring imputation for this variable are usually handled in the first attempt to find a donor, due to the mild set of constraints and large domain. The only imputation-revised variable created in the hot-deck step was IRMARIT.

4.3.2 Race Variables (Imputation Set 2)

As discussed in Section 4.2.7, race and Hispanicity were closely related. Therefore, race was used in the imputation of Hispanic/Latino origin and Hispanic/Latino group, and Hispanicity was used in the imputation of race. Since race underwent imputation first, imputation-revised versions of the Hispanic/Latino indicator and the Hispanic/Latino group were not available. This precluded their usage in race models. However, they were used extensively in constraints in the hot-deck step. The RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to 25, and 26 or older.

The race questions had low response rates relative to other questions in the NSDUH, due to high item nonresponse among Hispanic/Latino respondents. Nearly all of the race nonrespondents reported being of Hispanic/Latino origin (Table 4.1). The likeness constraints involving Hispanic/Latino group strongly influenced the final imputed values.

4.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain for the RP models included all unit respondents. Item respondents were those with EDRACE values from 1 to 15 and EDNWRACE values from 1 to 15. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.2.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the race variables were modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The outcome variable was the five-level variable EDRACEFORMODEL, which had the following levels:

1. White Only
2. Black/African American Only
3. American Indian/Alaska Native Only
4. Asian/Other Pacific Islander Only
5. Multiple Race

In survey years prior to 2008, multiple-race respondents were assigned to one of the first four categories above. An edited variable that did not include a category for more than one race was useful in the past because (1) the multiple race cell contained a small number of respondents, making imputation models difficult to fit; and (2) it was necessary to be used as a base variable for the final imputation-revised variable that did not include a category for more than one race (between 2003 and 2007, called IRRACE2). On the first point, the multiple racial category has become less sparse over time (refer to Section 3.3 of the 2008 imputation report [Ault et al., 2010] for more details). On the second point, because multiple-race respondents were classified as a separate category starting in 2008, a decision was made to cease to create IRRACE2, where multiple-race respondents were assigned a single race as shown in the first four categories above. It was replaced in most cases with the variable RACE4. The variable RACE4 is described in Section 4.3.3.4.

EDRACEFORMODEL is a recode of the variable EDRACE, described in Section 4.2.7.3.2:

EDRACEFORMODEL =

- EDRACE, if $1 \leq$ EDRACE ≤ 4; else
- 5, if $5 \leq$ EDRACE ≤ 16; else
- missing.

The five predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDRACEFORMODEL.

4.3.2.3 Hot-Deck Step

Each item nonrespondent in the hot-deck step was assigned one of 11 missingness patterns. Ten of the missingness patterns, all rare, were set up to handle cases where something was known about the race categories such as "known to be Asian." The 11th missingness pattern, by far the most common, handled cases where nothing was known about the race categories. For a description of these missingness patterns, see Table E.5. Logical constraints applied to the

cases where something was known about the race categories. Otherwise, for the cases where nothing was known, only likeness constraints were used. Sometimes, what was "known" about the race categories came from a random assignment for indirectly mapped codes, as described in Section 4.2.7.2.1.

Besides the segment and delta likeness constraints, the likeness constraints based on Hispanic/Latino group were important determinants of the final imputed value, because the vast majority of the item nonrespondents for race were Hispanic/Latino. In 2012, 16.81 percent of the overall respondent pool were of Hispanic/Latino origin, but 97.91 percent of the item nonrespondents for race were of Hispanic/Latino origin. Table 4.1 reports the distribution of Hispanic/Latino group among race item nonrespondents in 2012. Almost all are Hispanic/Latino and most (more than two thirds) of the Hispanic/Latino nonrespondents are Mexican only.

Table 4.1 Hispanic/Latino Status of Item Nonrespondents for Race

Hispanic/Latino Status	Item Nonrespondents for Race	
	Number	**Percentage**
Not Hispanic/Latino or Missing Hispanic/Latino Indicator	50	2.09
Hispanic/Latino	2,345	97.91
Mexican Only	1,677	70.02
Puerto Rican Only	236	9.85
Central/South American Only	185	7.72
Dominican Only	114	4.76
Other/Unknown	133	5.55
Total	2,395	100.00

Depending on the missingness pattern, the item nonrespondent received values from the selected donor for some subset of EDRACE, EDNWRACE, EDQD051-EDQD055, and a collapsed version of EDQD056-EDQD0513. The collapsed version of EDQD056-EDQD0513 is simply the minimum of these variables; this is an indicator of whether the respondent was Asian. Most receive values for all variables. Item nonrespondents in missingness pattern 2 (known to be Asian but missing an Asian finer category) received values for only EDNWRACE, and item nonrespondents in missingness pattern 3 (known to be multiple race, but no other information) receive values for everything except EDNWRACE. The imputation-revised versions of these variables are shown in Table 4.2.

Table 4.2 Edited Race Variables and Their Imputation-Revised Counterparts

Edited Race Variable	Imputation-Revised Race Variable
EDQD051	IRRACEWH
EDQD052	IRRACEBK
EDQD053	IRRACENA
EDQD054	IRRACENH

Table 4.2 Edited Race Variables and Their Imputation-Revised Counterparts (continued)

Edited Race Variable	Imputation-Revised Race Variable
EDQD055	IRRACEPI
EDQD056-EDQD0513 (collapsed)	IRRACEAS
EDRACE	IRDETAILEDRACE
EDNWRACE	IRNWRACE

IRDETAILEDRACE is not included in the final data files because the information it contains is covered by the other imputation-revised race variables. It is used in a likeness constraint for the Hispanic/Latino group variable discussed in Section 4.3.4.

Due to the strict constraints, the delta constraint had to be dropped sometimes. However, the likeness constraints related to Hispanic/Latino group were never dropped.

4.3.3 Hispanic/Latino Origin Variable (Imputation Set 3)

For the Hispanic/Latino origin indicator, the RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to 25, and 26 or older. Details on the procedures are given in the next four sections. The base variable for imputation, EDHOIND, is described in Section 4.2.7.4.1. The item response rate for this variable was much higher than for race.

4.3.3.1 Response Propensity Step

The response propensity models for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for EDHOIND. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.3.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the Hispanic/Latino origin indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.[48] The outcome variable was EDHOIND. The single predicted mean was the predicted probability that the respondent was of Hispanic/Latino origin.

4.3.3.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino origin indicator included a single predicted mean from the PRD step, no logical constraints, and only the segment and delta likeness constraints. EDHOIND is the base variable for imputation, and the imputation-revised version is called IRHOIND. Details on the hot-deck step, including the likeness constraints, are available in Tables E.6 and E.7 in Appendix E.

[48] In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

4.3.3.4 Recodes for Additional Race/Ethnicity Variables

The imputation-revised race (IRNWRACE) and imputation-revised Hispanic/Latino indicator (IRHOIND) variables were used to create several additional combined race/ethnicity variables. One of these (RACE4) was used in the subsequent processing of imputation-revised variables and had four levels: non-Hispanic/Latino white, non-Hispanic/Latino black/African American, Hispanic/Latino, and non-Hispanic/Latino other/multiple race. The NEWRACE1 and NEWRACE2 variables also were created from IRNWRACE and IRHOIND and were used extensively in the production of the 2012 detailed tables (Center for Behavioral Health Statistics and Quality, 2013a).

4.3.4 Hispanic/Latino Group Variable (Imputation Set 4)

The edited variable EDHOGRP, described in Section 4.2.7.4.3, categorized Hispanic/Latino respondents into Hispanic/Latino groups. These categories were directly mapped to the same categories in the imputation-revised variable, IRHOGRP4, which had eight possible values: Puerto Rican, Mexican, Cuban, Central or South American, Dominican, Spanish, Other Hispanic/Latino, and not Hispanic/Latino. The closely-related imputation-revised variable IRHOGRPM was also created to identify respondents who selected more than one Hispanic/Latino group; recall that a priority rule is used to assign a single group to multiple-group respondents in the creation of EDHOGRP (and therefore IRHOGRP4).

Imputations were not conducted separately within age groups, as was the case for marital status. The Hispanic/Latino group variables were created only for respondents of Hispanic/Latino origin as defined by IRHOIND. This results in a small domain. The models were likely to be better when age groups were combined because (1) none of the response categories were sparsely populated; and (2) sufficiently large donor pools were ensured in the hot-deck step.

4.3.4.1 Response Propensity Step

The response propensity models for imputation set 4 utilized the preliminary analysis weight, PANALWT. The domain indicator included all respondents of Hispanic/Latino origin as defined by IRHOIND. Item respondents were those with a nonmissing value for EDHOGRP who selected only a single Hispanic/Latino group. The multiple-group respondents whose EDHOGRP was assigned by the priority rule (Scenario 3, described in Section 4.2.7.4.3) were not used to fit the PRD model in the next step. See Table D.1 in Appendix D for details of the covariates used in the RP models for these variables.

4.3.4.2 Prediction Step

Because the model would have been much more difficult to fit if all seven levels were used, the EDHOGRP variable was collapsed into a four-level categorical variable (EDHOGRP2). Table 4.3 shows the mapping of EDHOGRP levels to EDHOGRP2 levels. Using the adjusted weights that are outputs of the RP step, EDHOGRP2 was then modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The four predicted means used in the subsequent hot-deck step were the predicted probabilities that the respondent had each value of EDHOGRP2.

Table 4.3 Mapping of EDHOGRP Levels to EDHOGRP2 Levels

EDHOGRP (Base Variable)	EDHOGRP2 (Modeled Variable)
Mexican	Mexican
Puerto Rican	Puerto Rican
Central or South American	Other Hispanic/Latino
Cuban	Cuban
Dominican	Other Hispanic/Latino
Spanish	Other Hispanic/Latino
Other Hispanic/Latino	Other Hispanic/Latino
Not Hispanic	Not Hispanic

4.3.4.3 Hot-Deck Step

The hot-deck step for the Hispanic/Latino group variables was straightforward. Besides the segment and delta likeness constraints, the most notable feature was a likeness constraint involving race. A five-level race variable was used as a covariate in the RP and PRD models with the following levels: White Only, Black/African American Only, American Indian/Alaska Native Only, Asian Only, and Multiple Race. To further exploit the relationship between race and Hispanic/Latino group, a likeness constraint required the donor's IRDETAILEDRACE variable to match a subset of the racial categories mentioned by the recipient. The constraint did not apply if the recipient was an item nonrespondent for race.

IRHOGRP4 was the imputation-revised version of EDHOGRP. The other imputation-revised variable IRHOGRPM was set equal to 8 (more than one Hispanic/Latino group) if either the respondent reported membership in more than one group, or the donor for a particular item nonrespondent reported membership in more than one group. Otherwise, IRHOGRPM was set equal to IRHOGRP4.

The Hispanic/Latino group variables generally have low imputation rates. The number of cases with missing data is usually fewer than 100 in each survey year.

4.3.4.4 Recodes for Additional Analyses

Among the recoded variables that were created from IRHOGRP4, the variable HISPGRP2 was used in subsequent processing and was created by collapsing the levels of IRHOGRP4 into four levels: Puerto Rican, Mexican, Other Hispanic/Latino (includes Cuban, Central or South American, Dominican, Spanish, and Other Hispanic/Latino), and not Hispanic/Latino.

4.3.5 Education Level Variable (Imputation Set 5)

The imputation-revised education level variable was similar to the imputation-revised Hispanic/Latino group variable in that it was categorical with numerous levels, and as with the Hispanic/Latino group, the response variable for the PRD model was collapsed into fewer levels for ease of modeling. There were generally very few missing cases for this variable—for some

years, fewer than 10—so the application of the method tended to be straightforward. Two age groups were used for RP and PRD modeling: 12 to 17 and 18 or older. However, the hot-deck step was implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older.

4.3.5.1 Response Propensity Step

The response propensity models for imputation set 5 utilized the preliminary analysis weight, PANALWT. The domain indicator for each of the two RP models included all unit respondents. Item respondents were those with a nonmissing value for EDEDUC. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

4.3.5.2 Prediction Step

EDEDUC was collapsed into fewer levels for modeling. The response variables were different for the two PRD models: the response variable for the 12-17 age group had five levels, and the one for the 18-or-older age group had four. The mapping of EDEDUC to the response variable RESPEDUC is shown in Table 4.4.

Table 4.4 Mapping of EDEDUC Levels to RESPEDUC Levels

EDEDUC (Base Variable)	RESPEDUC (Modeled Variable)	
	12-17	18+
Never attended school	Less than elementary school	Less than high school
1st grade completed		
2nd grade completed		
3rd grade completed		
4th grade completed		
5th grade completed		
6th grade completed	Elementary school	
7th grade completed		
8th grade completed	Middle school	
9th grade completed		
10th grade completed	Some high school	
11th grade completed		
12th grade completed	High school	High school
College or university/1st year completed		Some college
College or university/2nd year completed		
College or university/3rd year completed		
College or university/4th year completed		College or higher
College or university/5th or higher year completed		

Using the adjusted weights that are outputs of the RP step, both PRD models were fit using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. The predicted means matched the levels of the response variable, so there were five

predicted means for the 12-17 hot-deck step and four for the 18-25 and 26-or-older hot-deck steps.

4.3.5.3 Hot-Deck Step

The hot-deck step for the education level variable was straightforward and implemented separately for three age groups: 12 to 17, 18 to 25, and 26 or older. The only base variable was EDEDUC, and the imputation-revised version was called IREDUC. Both variables are based on the detailed 18-level variable, as compared with the simplified RESPEDUC variable used in the RP and PRD steps. No logical constraints were required. In addition to the segment and delta likeness constraints, the third likeness constraint required the donor to be the same age as the recipient. This was an especially important constraint for the 12-17 age group, because the age covariate often had to be dropped from the PRD model due to near-empty cells when the variables were cross-tabulated, causing instability in the estimates.

4.3.5.4 Recode for Additional Education Variable

EDUCCAT2, a recoded education variable, was created using the imputation-revised highest grade completed variable (IREDUC). EDUCCAT2 had five levels (less than high school and aged 18 or older, high school graduate and aged 18 or older, some college and aged 18 or older, college graduate and aged 18 or older, or 12 to 17 years old). This variable was often used as a covariate in later imputation models.

5. Editing and Imputation for the NSDUH Noncore Demographics Variables

5.1 Introduction

This chapter discusses procedures for editing and imputing the demographic variables from the "noncore" section of the interview. As noted in Chapter 1, demographic variables in noncore modules could be subject to change. The core demographics variables that were edited and imputed in the 2012 National Survey on Drug Use and Health (NSDUH) were discussed in Chapter 4.

This chapter discusses editing procedures for the following noncore demographics sections:

- moves in the past 12 months and State residency in the past 2 years,
- immigrant status,
- noncore education (i.e., education-related questions other than the highest grade attained),
- employment and workplace, and
- field interviewer (FI) debriefing questions (completed by the FI after the conclusion of the interview).

This chapter also discusses imputation procedures for the variables pertaining to immigrant status and current employment status. Other edited variables for noncore demographics that are discussed in this chapter were not imputed.

5.2 Editing the Noncore Demographics Variables

This section documents the editing procedures for the noncore demographics sections that were listed in Section 5.1. As noted in Section 1.1, the noncore demographics sections of the interview were administered by the interviewers using computer-assisted personal interviewing (CAPI).

5.2.1 Moves in the Past 12 Months and State Residency in the Past 2 Years

This section covers issues related to changes of residence. Specifically, question QD13 (edited variable MOVESPYR) asked respondents to report the number of times that they had moved in the past 12 months. If a respondent moved at least once in the past 12 months, the respondent was asked to report the State where he or she was living a year prior to his or her interview date (question QD13A; edited variable LIVSTYRA). If a respondent reported in the social environment module (for adults aged 18 older) or the youth experiences module (for youths aged 12 to 17) that he or she moved at least once in the past 5 years, the respondent was asked to report the State where he or she was living 2 years prior to the interview date (question

QD13B; edited variable LIVST2YA). Respondents who moved from another State to their current one (edited variable STATELOC; see Section 4.2.3) within the past 2 years were asked to report the month and year when they moved to their current State (question QD13C; edited variables MOVSTTMO and MOVSTTYR).

As noted in Section 4.2.3, the State that interviewers reported for the location of the dwelling unit sometimes did not match the State residence information that was used to sample a given case. If STATELOC was set to bad data because of incorrect information for the current State residency, MOVSTTYR and MOVSTMO also were set to bad data; respondents were asked question QD13C only if they previously reported in QD13A or QD13B that they had moved from a different State. Therefore, when STATELOC was set to bad data, this called into question whether respondents should have been asked to report the month and year when they moved to their current State.

As noted previously, information from question QD13C was captured in two edited variables pertaining to the month (MOVSTTMO) and year (MOVSTTYR) when respondents moved to their current State. Assignment of values to these two variables took into account the logic that the computer-assisted interviewing (CAI) program used to assign specific months and years to the response categories in QD13C, which depended on the interview date and the answers in questions QD13A and QD13B. Specifically, if QD13A had been answered (i.e., for the State where the respondent was living a year ago) and the State where the respondent lived a year ago did not equal the respondent's current State residence, then the respondent saw the response categories that were filled in QD13C as follows: the first response option (QD13C = 1) was filled with the month and year that occurred 12 months ago (i.e., same month as the interview month but in the year 2010), the next response option was filled with the month that occurred 11 months ago, and so on, up through the last response option (QD13C = 13), which was the current month in which the respondent was interviewed.

Suppose, for example (not necessarily actual data), that a respondent was interviewed in May 2012, QD13A was answered, and QD13A was different from the State where the respondent currently was living. The first response option in QD13C that the respondent saw would have been "May 2011," corresponding to the same month as the interview month but in the prior calendar year. If a value of 4 was keyed in QD13C, that response corresponded to "August 2011" based on this interview month. In this situation, the edited variable MOVSTTYR was coded as 2011 and MOVSTTMO was coded as 8.

Otherwise, the month that the CAI logic filled in the response options for QD13C still began with the interview month. However, the year that was filled in the response options began with 2010. Response option 13 in QD13C was filled with the interview month in 2011. Suppose, for example, that QD13A was blank because the respondent did not move in the past 12 months (QD13 = 0), but the State that the respondent lived in 2 years ago (from QD13B) did not equal the State where the sample dwelling unit was located. If the interview was conducted in April 2012, then the first response option in QD13C would be filled with April 2010. If a value of 2 was keyed in QD13C, that response would correspond to the respondent moving to the current State in May 2010. In this example, MOVSTTYR was coded as 2010 and MOVSTTMO was coded as 5.

A key aspect of processing the variables LIVSTYRA, LIVST2YA, MOVSTTMO, and MOVSTTYR involved assignment of legitimate skip codes based on the CAI logic (Section 2.4.2). This assignment of legitimate skip codes occurred as follows:

- If respondents did not move in the past 12 months (MOVESPYR = 0), then LIVSTYRA was assigned a legitimate skip code.
- If respondents did not move in the past 5 years (i.e., SNMOV5YR = 0 from the social environment module for respondents aged 18 or older or YEMOV5YR = 0 from the youth experiences module for respondents aged 12 to 17), then LIVST2YA was assigned a legitimate skip code.
- If respondents moved zero times or moved within the same State, then MOVSTTYR and MOVSTTMO were assigned legitimate skip codes. This condition would not hold if respondents reported in either QD13A or QD13B that they moved to their current State from outside of the United States.

The following additional issues were relevant to editing of LIVSTYRA, LIVST2YA, MOVSTTMO, and MOVSTTYR:

- Question QD13 for the number of moves in the past 12 months was answered as "don't know" (DK) or "refused" (REF). However, the respondent indicated no moves in the past 5 years in the social environment module (for adults; edited variable SNMOV5YR) or the youth experiences module (for adolescents; edited variable YEMOV5YR). In this situation, it could be logically inferred that the respondent also did not have any moves in the past 12 months. A value of 0 was logically assigned to MOVESPYR (corresponding to QD13). An "editing indicator" variable EIMOVPYR was created to indicate when this logical inference was made. The default value for EIMOVPYR was 1, when no editing was done to MOVESPYR. EIMOVPYR was assigned a value of 2 when this edit was implemented.
- The number of moves in the past 12 months (from MOVESPYR) was greater than the number of moves in the past 5 years from SNMOV5YR or YEMOV5YR. No editing was done to MOVESPYR, SNMOV5YR, or YEMOV5YR. However, a "flag" variable (MOVYRFLG) was created that indicated when this inconsistency occurred. The default value was 98 (i.e., blank) when MOVESPYR was consistent with either SNMOV5YR (for adults) or YEMOV5YR (for adolescents). MOVYRFLG had a value of 1 when MOVESPYR was inconsistent with the SNMOV5YR or YEMOV5YR.
- Question QD13B (edited variable LIVST2YA) was skipped because respondents reported no moves in the past 5 years in SNMOV5YR or YEMOV5YR. However, MOVESPYR indicated one or more moves in the past 12 months. In this situation, respondents logically should have reported that they had no moves in the past 12 months if they previously reported no moves in the past 5 years. Therefore, LIVST2YA retained a code of 98 (i.e., blank) rather than being assigned a legitimate skip code, even though SNMOV5YR or YEMOV5YR indicated no moves in the past 5 years.

5.2.2 Immigrant Status

Edits described in this section for the immigrant status variables have applied since 2004, when the content of the immigrant status variables changed. Question QD14 asked whether respondents were born in the United States. If they were not born in the United States, question QD15 asked respondents for their country of birth. The former question QD16 (length of time that respondents had lived in the United States, corresponding to edited variable LIVEDUSA) was replaced in 2004 with three variables pertaining to the following:

- whether respondents had lived in the United States for at least 1 year (question QD16A or edited variable LIVUS1YR);
- the number of years that respondents had lived in the United States, if they reported in QD16A that they had lived in the United States for at least 1 year (question QD16B or edited variable LIVUSYRS); and
- the number of months that respondents lived in the United States, if they answered QD16A as "no," indicating that they had not lived in the United States for at least 1 year (question QD16C or edited variable LIVUSMOS).

An important aspect of processing the immigrant status variables involved assigning legitimate skip codes where relevant (Section 2.4.2). For example, if respondents reported that they were born in the United States (i.e., the edited variable BORNINUS was answered as "yes"), the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS were assigned legitimate skip codes. Similarly, if LIVUS1YR = 1 (i.e., "yes"), LIVUSMOS was assigned a legitimate skip code. If LIVUS1YR = 2 (i.e., "no"), LIVUSYRS was assigned a legitimate skip code. When LIVUS1YR was coded as 94 ("don't know") or 97 ("refused"), the appropriate code for "don't know" or "refused" was assigned to the variables LIVUSYRS and LIVUSMOS that had been skipped.

If respondents reported that they were born outside the United States, however, it was possible for them to specify an answer in question QD15 that logically would mean that they were born in the United States. If this inconsistency occurred in the data (i.e., it had not been resolved by the interviewer), then the edited variable BORNINUS was logically inferred to be answered as "yes."[49] The edit procedures also logically inferred that the edited variables BORNINOT, LIVUS1YR, LIVUSYRS, and LIVUSMOS should have been skipped.

It also was possible for respondents to report in question QD16B that they had lived in the United States for a number of years greater than their current age. When this situation occurred, the edited variable LIVUSYRS was assigned a bad data code to indicate that the answer was inconsistent with the respondent's age. No editing was done to LIVUSYRS when LIVUSYRS was equal to the respondent's age because that answer in LIVUSYRS could be interpreted to mean that the respondent came to the United States as a baby.

Because an imputation-revised variable was desired for the age of entry to the United States for respondents who were born elsewhere, the final step in editing these variables was to

[49] If respondents reported being born in Alaska or Hawaii and were born before 1959 (i.e., before Alaska and Hawaii became States), these respondents still were considered to have been born in the United States.

use LIVUS1YR, LIVUSYRS, and LIVUSMOS to compute the age at which a respondent entered the United States. The age of entry to the United States was computed from continuous forms of the length of time that the respondent was living in the United States and the respondent's age. Because QD16B and QD16C were mutually exclusive, the edited variables LIVUSMOS and LIVUSYRS were used to create a continuous variable (LNGTHLIV) that indicated how many years a respondent had lived in the United States. In most cases, LNGTHLIV had the same value as LIVUSYRS. However, if the respondent had lived in the United States for less than 1 year, his or her LNGTHLIV values were obtained from LIVUSMOS by converting the number of months into fractions of 1 year. LNGTHLIV was set to missing when LIVUSYRS and LIVUSMOS had missing data codes. A continuous age variable, CONTAGE, was defined as CONTAGE = (interview date – birth date + 1)/365.25. Because the interview date and birth date, as described in Chapter 4, had no missing values, CONTAGE also had no missing values. For respondents who were born in the United States, a legitimate skip code of 999 was assigned to both the LNGTHLIV and CONTAGE variables.

The variable ENTRYAG2 was the base variable for creating the imputation-revised variable IRENTAG2 and represented the (continuous) age at which an immigrant entered the United States. ENTRYAG2 was defined as ENTRYAG2 = CONTAGE – LNGTHLIV and was set to missing if LNGTHLIV was missing. ENTRYAG2 also had a legitimate skip code (999) for respondents who were born in the United States.

5.2.3 Noncore Education

The noncore education module assessed whether respondents were currently enrolled in school and, if not, whether respondents who were aged 12 to 25 had dropped out of school without having received a high school diploma. Figure 5.1 shows the logic for asking follow-up questions according to whether or not respondents were enrolled in school. Respondents who reported that they were enrolled were asked to report their current grade in school (or the grade they would be in once they returned from school break), whether they were a full-time or part-time student, and if they were full-time students, the number of days that they missed school in the past 30 days because they were sick or because they skipped school (questions QD18 through QD21). Respondents who were aged 25 or younger, had completed the 12th grade or lower (from question QD11), and were not enrolled in school were asked whether they had received a high school diploma (question QD22). Respondents in this age group who reported that they left school without receiving a high school diploma were asked whether they had received a general educational development (GED) certificate of high school completion, why they left school before receiving a high school diploma, and their age when they left school (questions QD23 through QD25). Table 5.1 shows the edited variables that corresponded to the questions in the noncore education module.

Consistent with this module being in a noncore section of the interview, the content of questions in this module has changed over time. However, these questions have not changed since the start of the new NSDUH baseline in 2002. Documentation of historical changes to the noncore education questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 Methodological Resource Book (MRB) (Kroutil, Chien, Handley, & Bradshaw, 2013).

Figure 5.1 Follow-Up Question Logic for Respondents in the Noncore Education Module

GED = general educational development.

Table 5.1 Mapping of Noncore Education Questions to Edited Variables

Question	Edited Variable	Question Text
QD17	SCHENRL	The next questions are about school. Are you now attending or are you currently enrolled in school? By "school," we mean an elementary school, a junior high or middle school, a high school, or a college or university. Please include home schooling as well.
QD17A	SCHENRL	Are you currently on a holiday or vacation break from school?
QD17B	SCHENRL	Do you plan to return from school when your holiday or vacation is over?
QD18	EDUCATND	What grade or school [are you now attending/will you be attending when your vacation is over]? Please tell me the number from the card.
QD19	SDNTFTPT	[Are you/Will you be] a full-time or a part-time student?
QD20	SCHDSICK	During the past 30 days, that is, from [DATEFILL] up to and including today, how may whole days of school did you miss because you were sick or injured?
QD21	SCHDSKIP	During the past 30 days, that is, from [DATEFILL] up to and including today, how may whole days of school did you miss because you skipped or "cut" do just didn't want to be there?
QD22	HSDIPLMA	Have you received a high school diploma?
QD23	HSGED	Have you received a GED certificate of high school completion?
QD24	LFSCHWHY	Please look at this card and tell me which one of these reasons best describes why you left school before receiving a high school diploma.
QD24SP	LFSCHWHY	What is the main reason why you left school before receiving a high school diploma?
QD25	LFTSCHAG	How old were you when you stopped attending school?

NOTE: Questions QD18 through QD21 applied if respondents reported in QD17 through QD17B that they were currently in school or planned to return to school when their break was over. Questions QD22 through QD25 applied to respondents aged 12 to 25 who reported that they were not currently in school or were on break but did not plan to return to school when their break was over. See Figure 5.1 for additional routing within the noncore education module.

The remainder of this section discusses the following issues that were relevant to editing of the noncore education variables:

- overall routing logic for the noncore education module (Section 5.2.3.1);
- edits involving respondents' school enrollment status (Section 5.2.3.2);
- general edits involving the last and current grade for respondents who were in school (Section 5.2.3.3);
- specific edits involving the last and current grade for respondents aged 12 to 18 who were in school (Section 5.2.3.4);
- specific edits involving the last and current grade for respondents aged 19 or older who were in school (Section 5.2.3.5); and
- miscellaneous edits involving the noncore education variables (Section 5.2.3.6).

Table B.1 in Appendix B describes edits for the noncore education variables pertaining to current school enrollment, last and current grade, and receipt of a high school diploma or GED.

Table B.1 also is intended as a companion to the discussion of the editing procedures for the noncore education variables in Sections 5.2.3.2 through 5.2.3.5.

5.2.3.1 Routing Logic for Noncore Education

As shown in Figure 5.1, determining whether respondents were currently attending school was critical for determining the subsequent questions that respondents would be asked about their education. However, respondents could answer the question about their current school attendance (QD17) as "no" if they were on a holiday break or vacation from school. Therefore, respondents who did not report in question QD17 that they were currently attending school were asked follow-up questions (if they were aged 12 to 25 and their highest reported grade from question QD11 was grade 1 to 15) to determine if they were on a holiday or vacation break from school (question QD17B), and if so, whether they intended to return to school once their break was over (question QD17B).

If respondents reported in QD17 that they were not currently attending school (QD17 = 2) but they were on break from school (QD17A = 1) and they intended to return to school once their vacation or break was over (QD17B = 1), the edited school enrollment variable SCHENRL (based on data from QD17, QD17A, and QD17B) was set to a value of 1 ("yes") to indicate that the respondents should be considered enrolled. Otherwise, the response from QD17 was carried over to SCHENRL. That included situations in which respondents reported in QD17A that they were not on vacation break from school, or who reported in QD17B that they were on break but did not intend to return to school once their break was over.

If respondents were currently attending school or were on break but intended to return, the wording of subsequent questions varied according to their current status. For question QD18, respondents who reported in QD17 that they were currently attending school (QD17 = 1) were asked to report their current grade in school (i.e., at the time of the interview). For respondents who were on vacation break from school but intended to return to school once their break was over (QD17B = 1), question QD18 asked for the grade that they would be in once they returned from their vacation break. Similarly, for the question about full-time or part-time student status (question QD19) respondents who were currently attending school were asked, "Are you a full-time student or a part-time student?" Those who were on break from school but intended to return to school were asked, "Will you be a full-time or a part-time student?"

Based on the logic shown in Figure 5.1, a key aspect of editing the noncore education variables involved assigning legitimate skip codes (Section 2.4.2) based on respondents' current enrollment status and age. If respondents were currently enrolled in school, the edited variables corresponding to questions QD22 through QD25 (HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG) were assigned legitimate skip codes. Similarly, respondents aged 26 or older were considered to have legitimately skipped out of questions QD22 through QD25 because of the age requirement for administration of these questions, regardless of whether they might not have finished high school. In addition, if respondents were not currently enrolled in school, the edited variables corresponding to questions QD18 through QD21 (EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes, provided there were no other data to suggest that they were enrolled (see Section 5.2.3.2).

5.2.3.2 Edits Involving Current School Enrollment

Because the routing logic in the noncore education module was based on respondents' current enrollment status, an important aspect of editing the variables in this module involved editing of the enrollment variable SCHENRL. Table B.1 in Appendix B describes edits for SCHENRL.

For example, respondents aged 12 to 25 who reported that they were not enrolled in school and that they had not received a high school diploma were asked to report why they left school before receiving a diploma (Figure 5.1). If respondents reported leaving school for some other reason besides the ones that were listed in question QD24, they were asked to specify the main reason why they left school. If the "OTHER, Specify" response (see Section 2.3.1.4) indicated that the respondent was still in school or was being home schooled, SCHENRL was assigned special codes to indicate that the respondent was in school (see Table B.1). Except for the data on reasons for leaving school that were responsible for the logical inference that the respondent was in school (edited variable LFSCHWHY), the variables HSDIPLMA and HSGED were overwritten with values of 89, and LFTSCHAGE was overwritten with a value of 989 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2) because it was logically inferred that these questions should have been skipped. Also, because questions QD18 through QD21 had been skipped based on the respondent not having reported current school enrollment, the edited variables corresponding to these questions retained missing values (i.e., codes of blank).

In addition, a "hard error" is included in the education section for situations in which the highest grade from QD11 was higher than the current (or anticipated) grade from QD18. FIs were prompted to verify the answers with the respondents and correct any information in QD11 or QD18. If the answers were correct as recorded, the FIs could "suppress" the hard error and continue with the interview. [50] When FIs suppressed a hard error message, however, they were requested to enter a comment documenting why the information that had been entered in QD11 and QD18 was correct. Although most of these comments were relevant to editing involving the last and current grade (Section 5.2.3.3), some comments were relevant to editing the school enrollment variable.

Specifically, if the FI's comments indicated that the respondent was now in some sort of technical or vocational school, the school enrollment variable SCHENRL was set to a value of 4 (No LOGICALLY ASSIGNED). This edit was done because interviewers were instructed not to include vocational or technical schools as types of schools in which respondents could be enrolled. When SCHENRL was set to a value of 4, any data in EDUCATND, STUDNT, SCHDSICK, and SCHDSKIP were overwritten with values of 89 (LEGITIMATE SKIP Logically assigned). Where possible, when respondents were inferred not to be enrolled in school because their current enrollment was in a technical or vocational school, FI comments

[50] Unlike the consistency checks that were described in Section 1.2 and elsewhere in this report (particularly Chapter 6), "hard errors" in the CAI instrument typically required interviewers or respondents to resolve an inconsistency before the interview could proceed to the next available question; consistency checks in self-administered sections of the interview gave respondents an opportunity to resolve an inconsistency but did not require resolution before the interview could proceed. As noted in Section 5.2.3.2, however, interviewers could allow an inconsistency to remain between the last and current grade but were required to provide a reason for doing so.

also were used to edit the variables pertaining to receipt of a high school diploma (HSDIPLMA) or receipt of a GED certificate of high school completion (HSGED). These edits are discussed further in Section 5.2.3.6.

5.2.3.3 General Edits Involving the Last and Current Grade

The current school grade from question QD18 could be inconsistent with the highest grade that the respondent reported completing in question QD11. In most situations, one might expect the current grade in QD18 to be one grade level higher than the response in QD11. In addition, no editing was done when the current grade reported in QD18 was the same as the highest grade reported in QD11 because respondents could have been repeating a grade.

As noted in Section 4.2.9, the core education variable EDUC (highest grade completed) was not edited with respect to data on a respondents' current grade because QD18 was in a noncore section of the interview. However, a second variable, EDTEDUC, was created as part of the editing of the noncore education module. Consequently, the core education variable would not be affected by changes that might occur in the content of noncore education variables over time. Nevertheless, the EDTEDUC variable might in some situations be a more accurate reflection of the highest grade that respondents had completed.

As noted previously, FIs could enter comments if they suppressed a hard error message when the highest grade from QD11 was higher than the current (or anticipated) grade from QD18. These comments were reviewed on a case-by-case basis to determine which of the following held:

- the answers should be accepted and no editing should be done to EDTEDUC (corresponding to QD11) or EDUCATND (corresponding to QD18);
- the value for EDTEDUC or EDUCATND should be edited for consistency with the comments entered by the FI;
- EDTEDUC or EDUCATND should be set to bad data based on the FI comments; or
- predefined editing rules for education should be invoked (see below and Table B.1).

Any edits based on the FI comments were done on a case-level basis using the respondent ID rather than on an automated basis using predefined editing rules. These case-level edits superseded any of the usual edits discussed in Table B.1 that otherwise would have been done. However, this hard error was suppressed for fewer than 30 cases, and specific case-level edits were done for fewer than 15 of these cases.

The general education edits discussed in the remainder of this section were invoked if the hard error between QD11 and Q18 had been triggered, the answers from QD11 and QD18 had not been corrected, or the FI's comments indicated that a correction needed to be made, but what needed to be corrected was not clear from the FI's comments. However, answers to QD11 and QD18 were accepted when FIs provided a plausible reason for the discrepancy between the two answers, such as if respondents were in college and transferred to another school but some prior credits did not transfer.

The following potential patterns of inconsistent or questionable data could occur between QD18 and QD11 despite the presence of the hard error check between the two questions:

- the hard error was triggered, but the case was allowed to proceed through the general education edits for the reasons described above;
- the hard error was not triggered, the current grade in QD18 was exactly two grades higher than the highest grade completed (from QD11), but the respondent was at a current grade level that would be expected for someone at his or her age (e.g., if a 12 year old reported last completing the 4th grade and reported currently being in the 6th grade); or
- the hard error was not triggered, and the current grade in QD18 was more than two grade levels higher than the highest grade from QD11.

An algorithm was developed to handle these types of situations when they occurred. This algorithm is discussed in detail below. In particular, having accurate data on respondents' current grade levels is important for comparing NSDUH data with drug use data from in-school surveys, such as Monitoring the Future, that are administered to students in specific grades.

For respondents aged 12 to 23, a series of arrays was set up that mapped out the highest grade and current grade that would be *expected* relative to a respondent's current age, assuming an orderly progression from one grade level to the next highest level. Table 5.2 shows a matrix that maps the current age with expected grades in the United States.

Table 5.2 Mapping of Current Age with Expected Grades

Current Age	12	13	14	15	16	17	18	19	20	21	22	23
Expected Completed Grade	6	7	8	9	10	11	12	13	14	15	16	17
Expected Current Grade	7	8	9	10	11	12	13	14	15	16	17	17

For example, one might expect most people in the United States to have completed the 6th grade by the time they are 12. It would therefore not be unreasonable for someone to be aged 12 and to be currently in the 7th grade, depending on when the respondent was interviewed. An upper age limit was set at 23 because a grade level of 17 (college or university, 5th year or higher) was the upper limit of the education levels.

In addition, the algorithm allowed for some deviation relative to the expected ages. For example, if a respondent was aged 12, had completed the 5th grade, and was currently in the 6th grade, this would be an acceptable pattern because the respondent might have had his or her 12th birthday at some point during the 6th grade.

Separate edits were done depending on whether a respondent was aged 12 to 18 or was older than 18. The rationale for doing edits separately for these two different age groups was that the typical progression from one grade level to the next would be less likely to hold for adults and at higher educational levels. Suppose, for example, that a respondent completed 3 years of college but changed majors and not all of the prior credits applied to the new major. It would be

possible for the respondent to report having completed 3 years of college and to be currently enrolled at a level lower than the third year of college, depending on how the respondent interpreted these questions. Similarly, a respondent who got a bachelor's degree in one field and went back to school for a second bachelor's degree might report having completed 4 years of college but also might report currently being enrolled at some level below the 4th year of college.

5.2.3.4 Edits Involving the Last and Current Grade for Respondents Aged 12 to 18

For respondents aged 12 to 18, the highest grade completed and the current grade were considered to be consistent with the respondent's age if what was reported was within 1 year of the grades given in Table 5.2. Thus, if a respondent was aged 12, the algorithm considered completion of any grades from the 5th through the 7th to be sufficiently consistent with the respondent's age. Similarly, for respondents aged 12, the algorithm considered current enrollment in any grades from the 6th through the 8th to be sufficiently consistent with the respondent's age.

Therefore, the following four data combinations were possible:

- both the completed grade and the current grade were consistent with the respondent's age;
- the highest completed grade was consistent with the respondent's age, but the current grade was not;
- the current grade was consistent with the respondent's age, but the highest completed grade was not; or
- neither the highest completed grade nor the current grade was consistent with the respondent's age.

Separate edits were done according to the four combinations of data patterns described immediately above. The following edits were done if both the completed grade and current grade appeared to be consistent with the respondent's age:

- If the current grade was more than two grade levels higher than the highest completed grade, the current grade was edited to be consistent with the highest grade because the latter was a core variable. For example, if a respondent was aged 17, reported completing the 10th grade, and reported currently being in the first year in college (QD18 = 13), the edits logically inferred that the respondent currently was in the 11th grade. The edited variable for current grade (EDUCATND) was assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).
- If the current grade was lower than the highest completed grade, the edit code gave precedence to the reported grade that would yield the most consistent result relative to the respondent's age. In particular, if accepting the report of the highest grade and inferring that the respondent was currently in the next highest grade would yield a current grade that was inconsistent with the respondent's age, then the noncore-

created variable EDTEDUC (i.e., edited highest grade completed) was assigned a value consistent with the current grade. Suppose, for example, that a 12-year-old respondent reported currently being in the 6th grade but completed the 7th grade. Accepting the answer that the respondent was currently in the 6th grade and had completed the 5th grade would be more consistent with the respondent's current age than would be the converse (i.e., accepting that this respondent had completed the 7th grade and inferring that he or she was currently in the 8th grade). In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If the highest completed grade was consistent with the respondent's age but the current grade was not, the highest completed grade was accepted by default. This was done if the current grade was lower than the highest completed grade or the current grade was more than two grade levels higher than the highest completed grade. The edited current grade EDUCATND was therefore assigned a value to indicate a current grade level that was 1 year higher than the highest completed grade. For example, if the respondent reported completing the 10th grade, EDUCATND would be assigned a code of 31 (i.e., 11th grade LOGICALLY ASSIGNED).

If the current grade was consistent with the respondent's age but the highest completed grade was not, the edit procedures accepted the current grade by default. Thus, if a 12-year-old respondent reported last completing the 4th grade and reported currently being in the 6th grade, this edit would identify the current 6th grade as being consistent with an age of 12; completing the 4th grade would not be identified as consistent with an age of 12. In this example, EDTEDUC would be assigned a code of 25 (i.e., 5th grade LOGICALLY ASSIGNED).

If neither the current grade nor the reported highest grade was consistent with the respondent's age, the following was done:

- If the current grade was lower than the highest grade that was reported, the algorithm picked the answer that was closest to the expected grade, based on the matrix shown in Table 5.2. The variable with the more inconsistent data was assigned a bad data code. This edit allowed for situations where respondents may have fallen behind where they would be expected to be grade-wise (e.g., if they had been held back a year).
- If the current grade was exactly two grade levels higher than the reported highest completed grade and the highest completed grade was higher than what would be expected for the respondent's age, no further editing was done. Otherwise, the created noncore variable EDTEDUC was assigned a bad data code. This edit was designed to allow for situations where a respondent might be on an accelerated track.
- If the current grade was more than two grade levels higher than the reported highest grade and it was lower than the expected current grade, then the value was retained for the current grade. The variable EDTEDUC was assigned a bad data code. In other situations, both EDTEDUC and EDUCATND (i.e., the edited current grade) were assigned codes of bad data. The rationale for the first edit was that, if EDUCATND was lower than the expected current grade, EDUCATND would be more consistent

with the *expected* current grade and the respondent's age than what the reported highest grade would be.

5.2.3.5 Edits Involving the Last and Current Grade for Respondents Aged 19 or Older

Minimal editing of EDTEDUC and EDUCATND was done for respondents aged 19 or older. Other than the edits described in this section, no other editing of the educational level data was done for respondents aged 19 or older.

If the current grade was lower than the highest completed grade and the current grade was at the 12th grade or lower, then EDUCATND (i.e., the current grade) was assigned a bad data code. Otherwise, no further editing was done when the current grade was lower than the highest grade. For example, if a respondent reported completing the 12th grade but the answer for the current grade indicated that the respondent was in the 1st grade (QD18 = 1), then the response in question QD18 would probably indicate a typographical error. The first edit described in this paragraph would assign a bad data code to EDUCATND.

If the current grade was more than two grade levels higher than the highest completed grade and the current grade was above the 12th grade, the edits compared what the highest grade completed would be relative to the current grade, if the highest completed grade were actually increased by 10 years. If increasing the highest completed grade by 10 years yielded a completed grade that was still less than or equal to the reported current grade, then the variable EDTEDUC was assigned a code of bad data. In this situation, the interpretation was that a typographical error was made for the highest grade. Otherwise, no further editing was done. The first edit described in this paragraph was based on observed patterns that suggested that keying errors may have been made in QD11 (highest grade completed). For example, there were respondents who reported completing the 1st grade (QD11 = 1) and currently being in their 13th or higher years of school. Again, this pattern suggested that the second digit did not get keyed in QD11. This edit ensured that the respondent was classified as being enrolled in a grade above the high school level.

5.2.3.6 Miscellaneous Edits Involving the Noncore Education Variables

Table B.1 discusses additional editing of variables in the noncore education module other than those pertaining to current enrollment and grade level. These included situations in which the data were consistent with respondents' status as being enrolled in school or not enrolled but other potential inconsistent responses were observed, such as responses for receipt of a high school diploma or GED (if not enrolled) or absences from school in the past 30 days (if currently enrolled).

For example, if a respondent reported that he or she was not enrolled in school, reported having received a high school diploma, but also reported in QD11 that he or she had completed only the 9th grade or lower, the respondent was logically inferred not to have received a diploma. HSDIPLMA was assigned a code of 4 (No LOGICALLY ASSIGNED).

In addition, where possible, when respondents were inferred not to be enrolled in school because of FI comments indicating that their current enrollment was in a technical or vocational

school, FI comments also were used to edit the variables pertaining to receipt of a high school diploma (HSDIPLMA) or receipt of a GED certificate of high school completion (HSGED). For example, if the FI comments indicated that respondents had received a high school diploma, HSDIPLMA could be assigned a code of 3 (Yes LOGICALLY ASSIGNED), and the remaining variables HSGED, LFSCHWHY, and LFTSCHAG could be assigned legitimate skip codes. In the absence of information in the FI comments that would permit editing of additional variables, HSDIPLMA, HSGED, LFSCHWHY, and LFTSCHAG were left as blank because these respondents who were logically inferred not to be enrolled were skipped out of questions that were relevant to respondents who were not enrolled.

5.2.4 Employment and Workplace

Respondents aged 15 or older were asked questions about their current employment, employment history, and characteristics of their workplace (if applicable). Question QD26 asked whether respondents worked in the week prior to the interview. If respondents reported that they did not work in the past week, they were asked in question QD27 whether they had a job or business. Respondents then were routed through different branches of work-related questions depending on how they answered these two key questions, as shown in Figure 5.2. For example, respondents who worked in the past week were asked questions to determine full-time or part-time work status (e.g., whether they usually worked 35 or more hours per week), whether they ever had a period of unemployment in the past 12 months, the number of days they missed work in the past 30 days because they were sick or because they did not want to be at their workplace, and characteristics of their workplace, particularly with respect to alcohol and other drug policies at their workplace. Similarly, respondents who did not work in the past week and did not have a job were routed into questions relevant for people who currently were not working, such as why they did not have a job, whether they made specific efforts to find work in the past 30 days, and the month and year when they last worked for pay.

The employment and workplace questions and logic underwent important changes prior to the start of the new NSDUH baseline in 2002. Documentation of historical changes to the employment and workplace questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 MRB (Kroutil et al., 2013).

In addition, coding procedures changed in 2003 for the industry in which respondents worked (currently or in the past year) and for respondents' occupations (for their current or previous job). Since 2003, the Census Bureau has coded these responses. In addition, the industry and occupation (I&O) codes and classification procedures since 2003 have been based on categories from the 2000 census. Therefore, the names of the I&O variables were changed, beginning with the 2003 NSDUH. The names of the I&O variables in 2002 were as follows: WRKINDUS (industry in which the respondent is currently working), WRKOCCUP (respondent's current occupation), WRKINDYR (industry in which the respondent formerly worked in the past year), and WRKOCCYR (respondent's former occupation in the past year). Since 2003, these variables have been called WRKIDSTY, WRKOCUPT, WRKIDSYR, and WRKOCUYR, respectively. Recoded versions of these variables (WRKIDST2, WRKOCUP2, WRKIDSY2, and WRKOCUY2) were created for the NSDUH public use file. Despite these changes, the procedures for editing these I&O variables have not changed.

Figure 5.2 Routing Logic for Respondents in the Employment Status Module

Also, since 2005, the question about the respondent's month of last employment (question QD39B) has been asked only for those respondents who reported in question QD39A (edited variable WRKLSTYR) that they last worked in the current year or the prior year (i.e., 2011 or 2012). Previously, all respondents who gave a valid year in question QD39A were asked to report the month they last worked in QD39B. Consequently, more respondents in 2005 legitimately skipped out of question QD39B compared with respondents in prior years. For this reason, the edited variable corresponding to question QD39A has been called WRKLSTMN since 2005. In 2004, this variable was called WRKLSTMO. If the year in WRKLSTYR was more than 1 year prior to the current survey year, WRKLSTMN was assigned a legitimate skip code. Because of the changes to the skip logic in 2005, the distribution of WRKLSTMN also was not comparable with the distribution of WRKLSTMO prior to 2005.

An important aspect of editing the work-related variables involved identification of situations where questions had been legitimately skipped (Section 2.4.2). A second key aspect of processing the work-related variables was to use the data to establish respondents' current work status (Figure 5.2). As noted previously, a single, recoded work status variable named JBSTATR was created that served as the starting point for creation of a simplified edited variable (EDEMPY) and a final, statistically imputed employment status variable (EMPSTATY). JBSTATR was created from the following final variables: WRKEDWK (whether the respondent worked in the past week), WRKHAVJB (whether the respondent had a job if he or she did not work in the past week), WRKHRSUS (whether the respondent usually worked 35 or more hours per week), WRKNOWRK (reason for not working in the past week despite having a job), WRKNOJOB (reason for not having a job in the past week), WRKEFFRT (made specific efforts to find work in the past 30 days), and WRKEDYR (whether the respondent had a job in the past 12 months). Based on the data in these variables, respondents aged 15 or older were assigned to one of the categories in JBSTATR that are listed in Table 5.3.

Table 5.3 Categories of JBSTATR

Code	Employment Situation	Code	Employment Situation
1	Worked at full-time job, past week	**12**	No job: in school/training
2	Worked at part-time job, past week	**13**	No job: retired
3	Has job but out: vacation/sick/temp absence	**14**	No job: disabled for work
4	Has job but out: layoff, looking for work	**15**	No job: didn't want a job
5	Has job but out: layoff, not looking for work	**190**	Has full-time job, reason for not working unknown
6	Has job but out: waiting to report to new job	**191**	Has part-time job, reason for not working unknown
7	Has job but out: self-employed, no business past week	**199**	Has job, no further information
8	Has job but out: in school/training	**290**	No job, no further information
9	No job: looking for work	**299**	Other, not in labor force
10	No job: layoff, not looking for work	Remaining codes in the 900 series have their standard meanings in NSDUH: Don't know (994), Refused (997), Blank (998), Legitimate skip (999)	
11	No job: keeping house full time		

In addition, respondents who reported in question QD31 that they did not have a job but were looking for work were not classified as being unemployed unless they reported in WRKEFFRT that they had made specific efforts in the past 30 days to find work (such as making contacts with someone about a job, sending out resumes or job applications, or placing or answering ads). If respondents reported that they did not have a job but were looking for work but WRKEFFRT was not answered as "yes," they were classified as not in the labor force (code 299) in JBSTATR.

If respondents did not know or refused to report whether they worked in the past week, WRKEDYR was checked for indications of whether respondents worked in the past year. Respondents who indicated in WRKEDYR that they did not work in the past 12 months were classified as not having a job (JBSTATR = 290). Otherwise, if respondents did not provide information on whether they worked in the past week (i.e., QD26 answered as "don't know" or "refused"), JBSTATR was assigned the corresponding code of "don't know" or "refused."

Table B.2 in Appendix B describes edits for employment variables pertaining to whether respondents had a job in the past week, the number of days absent from work in the past 30 days, the number of weeks without a job in the past 12 months, the number of hours worked in the past week, the year and month that respondents last worked for pay, and self-employment. As noted previously, for example, the question pertaining to the month that respondents last worked for pay was changed in 2005. If respondents reported in question QD39A that they never worked for pay, interviewers were instructed to enter a response of 9991. When the month question QD39B had been skipped because a response of 9991 had been entered in QD39A, the edited month variable WRKLSTMN was assigned a code of 91. Documentation for 9991 (or 91) was as follows: 9991 = NEVER WORKED AT A JOB OR BUSINESS.

A refinement to the editing procedures for the employment and workplace section also has been implemented since 2003 for respondents who reported that they did not work in the past week (WRKEDWK = 2). The variable pertaining to the number of hours that respondents worked in the past week (WRKHRSWK) was assigned a legitimate skip code regardless of how respondents answered the question about having a job (QD27). Prior to 2003, a legitimate skip code was assigned to WRKHRSWK only if respondents reported that they did not work in the past week (WRKEDWK = 2) and QD27 was answered as "no"; the prior logic did not assign legitimate skip codes to WRKHRSWK if respondents answered QD27 as "don't know" or "refused." Logically, however, if respondents reported that they did not work in the past week, they would not have worked any hours at a job during that period, regardless of how they answered question QD27.

A final procedure that is discussed for the editing of the employment variables (not presented in Table B.2) concerns creation of the base variable EDEMPY, which was used to create the imputation-revised employment status variable. EDEMPY was derived from JBSTATR and WRKHRSUS. WRKHRSUS was used in some cases to determine whether employed respondents were employed full time or part time. Specifically, EDEMPY was derived in the following manner:

EDEMPY =

- 99, if the respondent is 12 to 14 years old; else
- 1 (full time), if JBSTATR = 1 or 190, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 1; else
- 2 (part time), if JBSTATR = 2 or 191, or if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS = 2; else
- 3 (unemployed), if JBSTATR = 4, 5, 9, or 10; else
- 4 (other), if JBSTATR = 11, 12, 13, 14, 15, 290, or 299; else
- 5 (part time or full time), if JBSTATR = 3, 6, 7, 8, or 199 and WRKHRSUS was missing (i.e., greater than 2); else
- missing.

5.2.5 Field Interviewer Debriefing Questions

The FI debriefing section was to be completed by the interviewer to obtain information about the potential quality of the interview. That included information about factors that might have affected the quality of the data, such as the degree of privacy in the interview setting. These questions were not to be read aloud to the respondent.

Only minimal processing was done to the data in this section. Specifically, unedited variables were replaced with final, mnemonic variable names (e.g., PRIVACY for the variable pertaining to the interviewer's indication of how private the interview was). Where relevant, variables also were assigned legitimate skip codes based on the routing logic in this section.

5.3 Imputation for Noncore Demographics Variables

For the noncore demographics module of the 2012 NSDUH, three imputation-revised variables were created from the base variables EDEMPY, BORNINUS, and ENTRYAG2: the first was an employment status variable, EMPSTATY;[51] the second, IRBORNUS, was an indicator of whether the respondent was born in the United States; and the third, IRENTAG2, recorded the age at which the respondent entered the United States. These three variables were processed in three separate, single-member imputation sets, using the single response propensity (RP)/single prediction (PRD) type of the predictive mean neighborhood (PMN) method described in Chapter 3. The ultimate goal in imputing values for missing data in these variables was to create a data file containing variables that would indicate whether respondents could be included in past year incidence analyses based on when they entered the United States. All three variables tend to have item response rates of more than 99 percent. See Table A.25 in Appendix A for details on the rates of missingness among these three noncore demographics variables.

[51] Unlike other imputation-revised variables, EMPSTATY was not preceded by an "IR" prefix. However, it was accompanied by imputation indicators that did have the requisite "II" prefix: II2EMSTY and IIEMPSTY.

5.3.1 Employment Status (Imputation Set 1)

The first noncore demographic variable that underwent imputation was the employment status variable. The imputation process was straightforward except for one feature: the handling of cases with EDEMPY = 5, where it was known that the respondents were employed but it was not known whether they were employed part time versus full time. These cases were handled in the hot-deck step in a separate missingness pattern, with a single logical constraint and a modified predictive mean vector. The final imputation-revised variable EMPSTATY had five levels: employed full time, employed part time, unemployed, other, and a skip code for respondents aged 12 to 14. Two age groups were used for RP and PRD modeling: 15 to 25 and 26 or older. The hot-deck step was implemented separately for three age groups: 15 to 17, 18 to 25, and 26 or older.

5.3.1.1 Response Propensity Step

The response propensity models for imputation set 1 utilized the preliminary analysis weight, PANALWT. For the first RP model, the domain included all unit respondents aged 15 to 25. For the second RP model, the domain included all respondents aged 26 or older. In both cases, item respondents were those with EDEMPY values of 1, 2, 3, or 4. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

5.3.1.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the employment status variable was modeled using polytomous logistic regression as implemented by the MULTILOG procedure in SUDAAN. For both age groups, the outcome variable had four levels, which mapped to the first four levels of EDEMPY.

5.3.1.3 Hot-Deck Step

The predicted means from the prediction step play a central role in the donor selection algorithm applied in the hot-deck step, but unlike the RP and PRD steps, the hot-deck steps were run separately within three age groups: 15 to 17, 18 to 25, and 26 or older. This was done to allow for parallel processing, which decreases the time required for implementation. Each item nonrespondent in the hot-deck step was assigned one of two missingness patterns. Item nonrespondents with a missing value for EDEMPY were handled in the first missingness pattern, which used the full predictive mean vector and no logical constraints. Item nonrespondents with EDEMPY = 5 were handled in the second missingness pattern, which applied a logical constraint to ensure that the donor was employed (either full time or part time). Also, conditional probabilities were used to take advantage of the partial information that was available. Instead of using the model's predicted probabilities directly, a single predicted mean was derived using a conditional probability, which was the probability that the respondent was employed full time, given that the respondent was employed. In addition to the segment and delta likeness constraints, a third likeness constraint, that donor's age must be within 5 years of recipient's age, was applied in the hot-deck step. See Appendix E for more details on missingness patterns and constraints for employment status.

5.3.1.4 Recodes for Additional Analyses

EMPSTAT4 was a direct recode of EMPSTATY and AGE. For respondents who were younger than 15 or older than 17, EMPSTAT4 and EMPSTATY were equivalent. For 15- to 17-year-olds, responses for EMPSTATY were overwritten with a code indicating that the respondent was too young to have his or her employment status recorded for the variable. This was the same code that was used for 12- to 14-year-olds for EMPSTATY (and EMPSTAT4).

5.3.2 Immigrant Status, Born-in-U.S. Indicator (Imputation Set 2)

The second noncore demographic variable that underwent imputation was the born-in-U.S. variable, BORNINUS. This was a dichotomous variable with very few missing responses. The RP, PRD, and hot-deck steps were all run separately within three age groups: 12 to 17, 18 to 25, and 26 or older. The imputation procedure was straightforward and is described in the next three sections.

5.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all unit respondents. Item respondents were those with a nonmissing value for BORNINUS. See Table D.1 in Appendix D for details of the covariates used in the RP models for this variable.

5.3.2.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the born-in-U.S. indicator was modeled using dichotomous logistic regression as implemented by the LOGISTIC procedure in SUDAAN.[52] The outcome variable was BORNINUS. The single predicted mean was the predicted probability that the respondent was born in the United States.

5.3.2.3 Hot-Deck Step

The hot-deck step for the born-in-U.S. indicator included a single predicted mean from the prediction step, no logical constraints, and only the segment and delta likeness constraints. BORNINUS was the base variable for imputation and the imputation-revised version was called IRBORNUS. Details on the hot-deck step are available in Appendix E.

5.3.3 Immigrant Status, Age of Entry (Imputation Set 3)

The age of entry variable was created only for respondents who were not born in the United States as defined by IRBORNUS. This results in a small domain. As a result, imputations were not conducted separately within age groups. The models were likely to be improved when age groups were combined because (1) none of the response categories were sparsely populated, leading to more robust models; and (2) sufficiently large donor pools were ensured in the hot-

[52] In SAS-callable SUDAAN, this is the RLOGIST procedure to avoid confusion with SAS's own LOGISTIC procedure.

deck step. Details on the procedures applied to the age of entry variable are explained in the next three sections.

5.3.3.1 Response Propensity Step

The response propensity model for imputation set 3 utilized the preliminary analysis weight, PANALWT. The domain indicator for the RP model included all respondents who were not born in the United States as defined by IRBORNUS. Item respondents were those domain members with a nonmissing value for ENTRYAG2. See Table D.1 in Appendix D for details of the covariates used in the RP model for this variable.

5.3.3.2 Prediction Step

Using the adjusted weights that are outputs of the RP step, the predicted mean for an immigrant's age of entry was estimated using a linear regression model, as implemented by the REGRESS procedure in SUDAAN. To control the upper and lower bounds of predicted means for age of entry, it was necessary to perform a logit transformation on the response variable. The response variable in the model was the immigrant age at entry as a proportion of the continuous version of current age CONTAGE, as described in Section 5.2.2. The expression of the proportion is $P_i = Y_i/N_i$, where $Y_i = Age\ at\ Entry_i$ and $N_i = Continuous\ Age_i$ (CONTAGE).

After the weight adjustment, the following empirical logit transformation was used as the response variable in a weighted linear univariate regression:

$$\log\left[\left(Y_i + 0.5\right)/\left(N_i - Y_i + 0.5\right)\right].$$

This transformation was nearly equivalent to the standard logit transformation,

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

which was not used because this transformation is unstable for respondents who entered the country at their current age (such that $P_i = 1$).

5.3.3.3 Hot-Deck Step

Two logical constraints were utilized in the hot-deck step for the age of entry variable. Both involved the respondent's age. One required that the donor's age of entry be less than or equal to the recipient's current age. The other required that the difference between the recipient's current age and the donor's age of entry be less than 1 year if the recipient lived in the United States for less than 1 year (as indicated by QD16A) or greater than 1 year if the recipient lived in the United States for more than 1 year. The only base variable was ENTRYAG2 and its imputation-revised counterpart was IRENTAG2. The segment and delta likeness constraints were applied in the hot-deck step. Details on the hot-deck step are available in Tables E.17 through E.19 in Appendix E.

6. Editing and Imputation for the NSDUH Core Drug Use Variables

6.1 Introduction

This chapter provides documentation of procedures for the editing and statistical imputation (subsequently referred to as "imputation") of core drug variables in the 2012 National Survey on Drug Use and Health (NSDUH).[53] As a prerequisite for reviewing this chapter, readers are encouraged to familiarize themselves with the content and terminology in Chapters 1 through 3 of this report because these three chapters provide an overview to the NSDUH interview and the general approach to editing and imputation of NSDUH data, with information on specific processes that are common across the survey in Chapters 2 and 3.

Consistent with prior years, the core drug use measures collected in the 2012 NSDUH included lifetime use (or nonuse); initiation of use (i.e., age at first use and for recent initiates, the year, and month of first use); most recent use; frequency of use in the past 12 months (for alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, inhalants, and prescription psychotherapeutic drugs); and frequency of use in the past 30 days (for tobacco products except for pipe tobacco, alcohol, marijuana, cocaine, crack cocaine, heroin, hallucinogens, and inhalants). Figure 6.1 provides an overview of the general routing logic for the core drug sections of the interview.[54] The implications of this general routing logic for editing and imputation are discussed in connection with the procedures that are described in the remaining sections of this chapter.

Variables corresponding to questions for these measures first underwent editing to (1) replace missing values with codes that indicated that the questions did not apply; (2) replace missing values with nonmissing values that could logically be inferred from other data; (3) make logical inferences when answers to related questions were inconsistent; and (4) identify inconsistent or ambiguous responses to be resolved through imputation. Edited variables for these measures then underwent imputation to replace missing values with nonmissing values or to replace ambiguous responses (e.g., use at some point in a respondent's lifetime but no definite period for most recent use) with specific ones.

6.1.1 Edited (but Not Imputed) Drug Use Variables

Some core drug use variables each year undergo editing but no further imputation. Most of these variables pertain to use of tobacco products. Editing procedures for these tobacco variables are discussed in Section 6.2.9. Also, the variable corresponding to the usual number of alcoholic beverages that respondents consumed on days when they used alcohol in the past 30

[53] A discussion of "core" and "noncore" sections of the NSDUH interview is provided in Chapter 1. Section 6.1.3 discusses an exception to editing and imputation involving core drug variables.

[54] For additional information about the content and logic in the core drug modules for 2012, see the computer-assisted interviewing (CAI) specifications for the 2012 NSDUH, which are available at http://www.samhsa.gov/data/. The CAI specifications can be accessed by clicking on "Methodology Reports" and then by clicking on "NSDUH 2012 Methodological Resource Book" on the Methodology Reports Web page.

days (corresponding to question AL07) is used in editing other variables that subsequently undergo imputation (e.g., frequency of consumption of five or more drinks on a single occasion in the past 30 days, also referred to as the frequency of "binge" alcohol use) but the variable corresponding to question AL07 itself is not included in the subsequent imputation procedures. Editing procedures for the variable corresponding to question AL07 are discussed in Section 6.2.4.7.

Figure 6.1 Routing Logic for Respondents in the Core Drug Modules

AFU = age at first use; MFU = month of first use; YFU = year of first use.

6.1.2 "Parent" and "Child" Drug Use Categories

In this chapter, some of the drug use measures refer to a general drug category (e.g., hallucinogens), and other measures refer to one or more subcategories within a general category (e.g., LSD, PCP, or Ecstasy as types of hallucinogens). These drug categories are described using the terms "parent drug" for the general drug category and "child drug" for the drug subcategory. Parent/child drug pairs often (but not always) occurred in modules that included multiple gate

questions.[55] For example, the hallucinogens module includes questions about lifetime use of specific hallucinogens to determine the respondent's overall lifetime use or nonuse of any hallucinogen. Included in these questions are specific gate questions for LSD (LS01A), PCP (LS01B), and Ecstasy (LS01F). However, parent/child drug pairs also could appear in separate modules, such as for any use of cocaine and specific use of crack cocaine (i.e., if respondents reported lifetime use of any form of cocaine). Table 6.1 shows the drugs with parent/child relationships and the data that were collected for them.

Table 6.1 Core Drugs in the 2012 NSDUH with a Parent/Child Relationship

Parent Drug	Child Drugs	Parent Data Collected	Child Data Collected	"Other" Lifetime Use Indicator[1]
Smokeless Tobacco	Snuff, Chewing Tobacco	None; measures for initiation, recency, and tobacco brands were created from data for the child drugs[2]	Initiation[3], recency, 30-day frequency, tobacco brands	No
Cocaine	Crack	Initiation, recency, 12-month frequency, 30-day frequency	Initiation, recency, 12-month frequency, 30-day frequency	No
Hallucinogens	LSD, PCP, Ecstasy	Initiation, recency, 12-month frequency, 30-day frequency	Initiation, recency	Yes
Pain Relievers	OxyContin®	Initiation, recency, 12-month frequency	Initiation, recency, 12-month frequency	Yes
Stimulants	Methamphetamine	Initiation, recency, 12-month frequency	Initiation, recency, 12-month frequency	Yes

[1] See Section 6.2.8.1.
[2] A 30-day frequency for any smokeless tobacco use could not be determined if respondents were users of both snuff and chewing tobacco in the past 30 days.
[3] Initiation refers to the age at first use and, if applicable, the year and month of first use.

6.1.3 Special Situation for Methamphetamine, Stimulants, and Psychotherapeutics

Questions on methamphetamine use in NSDUH are first asked in the stimulants module in the core section of the questionnaire in the context of questions about nonmedical use of prescription stimulants.[56] One concern in measuring methamphetamine use in NSDUH is that some methamphetamine users—particularly those who used it in the past 12 months—may fail to report use if they do not recognize the drug when it is presented in the prescription drug context.

[55] See Section 2.2.3 for a discussion of the gate question terminology.
[56] Question ST01 in the core stimulants module asks, "Have you ever, even once, used Methamphetamine, Desoxyn, or Methedrine that was not prescribed for you or that you took only for the experience or feeling it caused? Methamphetamine also is known as crank, crystal, ice, or speed." Desoxyn® is available by prescription in the United States but is not commonly prescribed. Methedrine® is no longer available but could be relevant for measuring nonmedical use at some point in a person's lifetime.

To address this concern, questions were added to the special drugs module in the noncore section of the NSDUH questionnaire beginning in 2005 to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Findings from the methamphetamine analysis report in the 2005 NSDUH Methodological Resource Book (MRB; Ruppenkamp, Davis, Kroutil, & Aldworth, 2006) suggested that estimates of methamphetamine use based only on core data could be lower than the true population prevalence. However, larger estimates of methamphetamine use based on both core and noncore answers could be a partial artifact of asking a second set of questions *only* from persons who did not report use the first time. Repeating questions for any drug only to those who did not report use the first time could artificially increase the positive responses. Doing so only for methamphetamine could result in a disproportionate reporting of that drug relative to the others in the survey. In addition, because the respondents reporting methamphetamine use in the noncore questions essentially had contradicted their prior responses, some may have made mistakes in answering the noncore questions. For these reasons, additional follow-up items have been included since the 2006 NSDUH to identify respondents who failed to report methamphetamine use in response to the earlier question in the core stimulants module because they may not have considered methamphetamine to be a prescription drug. Respondents who did not previously report methamphetamine use because they did not consider it to be a prescription drug have been counted in core-plus-noncore (CPN) estimates of methamphetamine use.

For the purpose of examining trends in nonmedical methamphetamine use, a Bernoulli stochastic imputation (BSI) procedure was used in conjunction with the predictive mean neighborhood (PMN) method (see Chapter 3) to generate comparable estimates for 2002 through 2005. In addition to CPN variables for methamphetamine, CPN variables were created in the 2012 NSDUH for nonmedical use of stimulants and nonmedical use of any psychotherapeutic drug. Section 6.2.6 discusses editing procedures to create the CPN data for methamphetamine and stimulants. Section 6.3.7 discusses procedures to create the final imputed CPN variables for methamphetamine, stimulants, and any psychotherapeutic drug.

6.2 Editing Procedures for the Core Drug Use Variables

This section discusses procedures for editing the core drug use variables for tobacco (i.e., cigarettes, snuff, chewing tobacco, cigars, and pipe tobacco), alcohol, marijuana, cocaine in any form, crack cocaine, heroin, hallucinogens, inhalants, and nonmedical use of prescription-type pain relievers, tranquilizers, stimulants, and sedatives. Edits of the core drug use variables encompassed the following key activities:

- edits of the lifetime use variables (i.e., gate questions), where respondents indicated whether they have ever used the drug of interest (Section 6.2.1);
- edits of the recency-of-use variables, where respondents who indicated lifetime use of the drug indicated when they last used that drug (Section 6.2.2);
- edits involving users of only "child" drugs (Section 6.2.3);

- edits of the 12-month and 30-day frequency variables, where respondents who indicated use of a drug in the 12 months or 30 days prior to the interview indicated the number of days they used that drug in the period of interest (Section 6.2.4);
- edits for incidence (i.e., initiation) variables, where respondents who indicated lifetime use of a drug indicated when they first used it (Section 6.2.5);
- edits for methamphetamine and stimulants to take into account data for methamphetamine from the noncore special drugs module (Section 6.2.6);
- edits for lifetime daily cigarette use (Section 6.2.7);
- edits that were applied as part of the imputation processing (Section 6.2.8); and
- edits for drug use variables that do not undergo imputation (Section 6.2.9).

In connection with each of these edits, the discussion in the remainder of Section 6.2 focuses on relevant issues or inconsistencies in the data that needed to be addressed through logical editing. As noted in Chapter 1, the skip logic in the computer-assisted interviewing (CAI) instrument limited the chances for respondents to be routed to questions where they could give answers that were inconsistent with their answers to previous questions. For example, if respondents reported never using marijuana in question MJ01, the CAI program did not ask additional questions that would presume use of marijuana at least once. However, the CAI program did not completely eliminate all opportunities for respondents to provide inconsistent answers. Nevertheless, most processing of the CAI data was relatively straightforward, and the issues discussed in this section were not widespread relative to the total number of 68,309 respondents in 2012.

6.2.1 Edits of Lifetime Use Variables

As discussed in connection with the usable case criteria (Section 2.2.3), respondents were asked first whether they had ever used a drug of interest. For hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives (hereafter, "hallucinogens through sedatives"), respondents were asked a series of questions to establish whether they had ever used one or more specific types of drugs within that category (e.g., LSD, PCP, peyote). Only those respondents who indicated lifetime use of that drug (or lifetime use of one or more specific drugs within the respective modules for hallucinogens through sedatives) were asked more detailed questions about that drug (including situations in which respondents initially refused to answer a question about their lifetime use of a drug but then changed their answer to "yes" on follow-up; see Sections 2.2.3 and 2.4.2). Unlike the first six substance use modules[57] (i.e., for tobacco products through heroin), however, there was no overall question about lifetime use or nonuse for the next six substance use modules, (i.e., hallucinogens through sedatives). Rather, lifetime use of any drug in the overall category for these latter six modules could be deduced from one or more affirmative answers regarding lifetime use of any of the specific drugs in that category. Because

[57] The CAI specifications show a single module for all tobacco products. Within this module, there are separate gate questions for cigarettes, snuff, chewing tobacco, cigars, and pipe tobacco. For pipe tobacco, there is only one question following the gate question for this tobacco product (i.e., any use in the past 30 days). In addition, the CAI specifications show a separate module for crack cocaine. As discussed in Section 6.2.1.1, routing of respondents to the crack cocaine module was dependent on whether they reported lifetime use of any form of cocaine.

the modules for hallucinogens through sedatives included a question about use of "any other" drug in that category, answers of "no" for lifetime use of all of the specific drugs in that category (i.e., including lifetime use of "any other") logically indicated that the respondent had never used any drugs in that category.

Processing of the gate variables established whether (1) respondents had used a drug of interest at least once, (2) they had never used the drug, or (3) lifetime use or nonuse of the drug could not be determined. In addition to answering these gate questions as "yes" or "no," respondents could answer them as "don't know" or "refused." As noted in Section 2.4.2, final responses of "don't know" and "refused" were treated by the CAI skip logic as equivalent to situations where respondents never used the drug of interest. For the hallucinogens through sedatives, the CAI skip logic treated combinations of responses of "no," "don't know," and "refused" to the questions about individual drugs in the absence of any affirmative response to these questions in the same manner as if the respondent had answered all of these questions negatively. In these situations where a gate question was answered as "don't know" or "refused," the respondent's lifetime use or nonuse status was treated as unknown because these responses did not provide conclusive information one way or the other.[58] Cases with unknown lifetime use/nonuse status were subsequently imputed to be lifetime users or nonusers, as described in Section 6.3.

This step of the editing procedures also involved assignment of "bad data" codes to lifetime drug use variables (i.e., equivalent to missing data) if potential patterned responses previously had been identified (Sections 2.3.2 and 2.4.3). For the core interview sections pertaining to snuff through heroin,[59] these edits involved assignment of codes for bad data for the lifetime drug question and the follow-up probe, if respondents previously had refused the lead question (e.g., questions AL01 and ALREF for alcohol). For hallucinogens through sedatives, these edits involved assignment of codes for bad data to all lifetime variables in that section (e.g., data in questions LS01A through LS01H for hallucinogens and any associated "OTHER, Specify" data for hallucinogens).

6.2.1.1 Creation of Lifetime Use Variables for Tobacco Products through Heroin

Creation of edited variables for lifetime use or nonuse of tobacco products through heroin typically was straightforward because there were overall questions about lifetime use or nonuse. In most instances, therefore, codes corresponding to respondents' original answers simply were assigned to the relevant variable (e.g., 1 = Yes; 2 = No). When respondents were routed to a follow-up question because they initially refused to answer the first question about lifetime use or nonuse (see Section 2.4.2 in Chapter 2), their response to this follow-up question was taken as the final response for the lifetime use variable.

The exception to these procedures involved lifetime use or nonuse of crack cocaine. Specifically, the logic to question CC01, regarding lifetime use of any form of cocaine, governed

[58] For multiple gate drugs, this included situations where respondents answered one or more of the gate questions as "don't know" or "refused" but did not answer any of the other gate questions affirmatively.

[59] Because of the requirement of the usable case criteria that respondents had to have defined data for lifetime use or nonuse of cigarettes (Section 2.2.3), assignment of the missing value of bad data for lifetime use of cigarettes would result in a case being reclassified as not usable.

whether respondents were asked questions about crack cocaine. Therefore, the lifetime crack use variable CRKEVER (corresponding to question CK01) was assigned a code of 91 (i.e., never used cocaine) when COCEVER (corresponding to CC01) was answered as "no." If respondents reported lifetime use of cocaine but answered question CK01 as "no," then CRKEVER was assigned a code of 2.

6.2.1.2 Creation of Lifetime Use Variables for Hallucinogens through Sedatives

For the variables in the hallucinogens through sedatives sections that indicated lifetime use or nonuse of specific drugs, if respondents answered all gate questions in the series as "no," the edit procedures assigned codes of 91 to the entire series of variables (e.g., LSD through any other hallucinogen).[60] This was done to indicate that the respondents had never used any of the drugs in that category. In contrast, a code of 2 (i.e., "no") in edited variables for lifetime use of specific drugs within a broader category had the following meaning:

- the respondent was a user of at least one drug in the category but had never used the specific drug of interest (e.g., if a respondent was a lifetime user of LSD but not PCP); or
- the respondent reported never using the particular drug of interest but answered other questions in the series as "don't know" or "refused" (e.g., if a respondent did not know whether he or she had ever used LSD but definitely knew that he or she had never used PCP).

Variables also were created to indicate whether respondents had ever used one or more drugs within the overall drug category (e.g., any inhalant). Summary variables for these drugs could take on the following values:

- 1 ("Yes"). The summary variable for a given category was coded as 1 (i.e., "yes") if respondents answered "yes" for lifetime use of at least one specific drug in the overall category. The summary variable also was coded as 1 if respondents initially refused to answer all lifetime questions within one of these sections and then reported on follow-up that they had used some drug in that category at least once. For example, if at least one affirmative answer appeared in questions about lifetime use of inhalants (including the follow-up question INREF) was answered as "yes," INHEVER was coded as 1.
- 91 ("Never Used"). This code of 91 could be assigned to the overall summary variable for one of two reasons: (1) if respondents answered all questions about lifetime use of individual drugs within the category as "no," or (2) if respondents were routed to a follow-up question because they refused to answer *all* questions about lifetime use of individual drugs within a category and then they reported that they never used that type of drug. As noted previously, this code of 91 also was assigned to all of the individual lifetime use variables within that section. For

[60] In the hallucinogens module, this included situations in which respondents initially refused to report whether they had ever used LSD, PCP, or Ecstasy but then changed their answer(s) to "no" on follow-up. Similarly, in the stimulants module, this included situations in which respondents initially refused to report whether they had ever used methamphetamine but changed their answers to "no" on follow-up.

example, if respondents initially refused to answer all questions about lifetime use of specific inhalants, they were asked whether they had ever used any type of inhalant. If respondents answered this follow-up question as "no," the summary variable INHEVER was assigned a code of 91, and all individual lifetime inhalant variables were assigned codes of 91 as well.

- 97 ("Refused"). The summary variable was assigned a code of 97 if respondents initially refused to answer *all* questions about lifetime use for the specific drugs within one of the sections for inhalants through sedatives and then continued on follow-up to refuse to indicate whether they had ever used a drug within that overall drug category, In this case, the edited lifetime use variables for the individual drugs within that module also retained codes of 97. For example, INHEVER was assigned a code of 97 if a respondent refused to answer questions IN01A through IN01L for inhalants and also refused to answer the follow-up question INREF.
- 98 ("Blank"). The summary variable was assigned a code of 98 if respondents had one or more answers of "don't know" or "refused" to questions about lifetime use of specific drugs within the category (i.e., other than the previous situation in which respondents refused all questions) and they had no reports of lifetime use of any of these drugs. In this situation, the code of 98 indicated that lifetime use or nonuse for the overall category was unknown, even if one or more questions about lifetime use of specific drugs was answered as "no." For example, INHEVER was assigned a code of 98 if some questions in IN01A through IN01L were answered as "no" but other questions were answered as "don't know" or "refused."

6.2.1.3 Editing of "OTHER, Specify" Drug Variables for Nonusers

If respondents had never used any of the drugs in a series of multiple gate questions, they would not have been routed to questions where they could specify the use of some other drug in that overall category. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with codes of 9991 to indicate that the questions were skipped because the respondents had never used that class of drugs.

Similarly, if respondents reported in the preceding multiple gate questions that they used at least one drug in the category but they never used some other drug besides the ones they were asked about, they were legitimately skipped out of the questions that asked them to specify what "other" drug they had used. In this situation, blank values in the unedited "OTHER, Specify" drug variables were replaced with codes of 9999 (i.e., legitimate skip).

Functionally, the codes of 9991 and 9999 both indicate that respondents legitimately skipped out of the "OTHER, Specify" questions in that drug's section. However, the 9991 codes provide for analysts the extra level of detail that the respondents were legitimately skipped out of these questions because they had never used anything within that category of drugs.

6.2.1.4 Use of "OTHER, Specify" Drug Data to Edit Lifetime Use Variables for Hallucinogens through Sedatives

Table B.3 in Appendix B describes edits that were relevant to the multiple gate variables for hallucinogens through sedatives based on "OTHER, Specify" data *within the same module* (see Sections 2.4.1 and 2.4.5 in Chapter 2). For example, if a respondent did not report lifetime use of a specific drug in a module (e.g., PCP) but specified using it as another drug in the category (e.g., "marijuana laced with PCP" as another hallucinogen), the respondent was logically inferred to be a lifetime user of that drug. The following code was assigned to the edited lifetime use variable for that drug: 3 = Yes LOGICALLY ASSIGNED.

This edit was especially relevant to LSD, PCP, Ecstasy, OxyContin®, and methamphetamine, where lifetime use or nonuse was imputed if respondents had missing data for the lifetime use questions. Consequently, no imputation would be needed for these drugs if a respondent reported not knowing whether he or she had used these drugs or refused to report use of these drugs when asked directly but the "OTHER, Specify" data in that module indicated use.

6.2.1.5 Editing Issues for Lifetime Psychotherapeutic Variables

In the sections for the psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives), respondents were asked if they had ever used any of the medications below the red line on that drug's pill card.[61] If respondents answered "yes," they were asked to indicate which of the drugs they had used. If the respondents answered "no," they were skipped out of these follow-up questions. Therefore, consistent with the procedures described at the beginning of Section 6.2.1.2, if respondents had never used *any* prescription-type psychotherapeutic medications in that category (i.e., in addition to never having used any of the medications below the red line on that drug's pill card), the edits assigned codes of 91 (i.e., "never used") to all of the specific drugs that were skipped (e.g., codeine through Ultram® in the PR04A series). In contrast, if respondents reported never using any of the medications below the red line on the pill card but they reported use of at least one other drug (or they answered at least one other gate question as "don't know" or "refused"), the skipped drug questions were assigned codes of 99 (i.e., legitimate skip).

Tables B.3 and B.4 in Appendix B describe edits that were relevant to the gate variables for the psychotherapeutic drugs. As noted in Section 6.2.1.4, codes of 3 (Yes = LOGICALLY ASSIGNED) could be assigned to gate variables if respondents did not report nonmedical use of specific psychotherapeutic drugs but they specified nonmedical use. If respondents answered the lead question about nonmedical use of drugs "below the red line" (e.g., question PR04 for pain relievers) as "no" but they specified nonmedical use of any of these drugs, it was inferred that this question should have been answered as "yes." Therefore, the edited variable (e.g., ANLCARD, corresponding to question PR04) was assigned a code of 3, in addition to specific drugs (e.g., morphine) being assigned this code.

---

[61] As an aid in answering the questions for psychotherapeutic drugs, respondents could look at printed "pill cards" that showed pictures of prescription drugs that were included in a given module. Pill cards included a thick red line that separated groups of drugs above and below the line. For example, question PR04 asked, "Please look at the pain relievers shown below the red line on Card A. Have you ever, even once, used any of these pain relievers when they were not prescribed for you or that you took only for the experience or feeling they caused?"

The following additional codes could be assigned to the lifetime nonmedical use variables for psychotherapeutics:

4 = No LOGICALLY ASSIGNED,

81 = NEVER USED [DRUG] Logically assigned.

Assignment of both of these codes applied to situations where respondents specified use of over-the-counter (OTC) medications despite being instructed not to report about use of OTCs.

A code of 4 was assigned to the lead question for nonmedical use of any other drug in that category (e.g., ANLNOLST, corresponding to question PR05 for pain relievers) when respondents answered "yes" to at least one other gate question in that section for nonmedical use of psychotherapeutics or if respondents had answers of "don't know" or "refused" in other gate questions. Codes in the "OTHER, Specify" variables were overwritten with codes of 9989 (LEGITIMATE SKIP Logically assigned; see Section 2.4.2) to indicate that respondents logically should have skipped these questions. Assignment of a code of 4 to the lead question for nonmedical use of any other drug in that category did not affect the editing of other gate questions in that section.

Codes of 81 were assigned when the respondent answered "no" to all of the questions about lifetime use of specific medications in that category except for use of any other medication in that category (e.g., any other pain reliever besides the ones shown on Pill Card A); and the *only* thing the respondent specified was an OTC medication, subject to the qualifications discussed in the remainder of this section.[62] If respondents who used only OTC drugs correctly followed instructions and answered all gate questions for a given psychotherapeutic category as "no," they would be skipped out of the remaining questions for that category of psychotherapeutic drugs (e.g., first use, most recent use). Consequently, the exclusive OTC users who incorrectly answered the question about nonmedical use of other drugs in the category as "yes" would comprise some unknown (and possibly unrepresentative) subset of exclusive OTC users.

Edits that assigned codes of 4 or 81 applied if the only responses in the "OTHER, Specify" data for a type of psychotherapeutic medication were OTCs, with the remaining "OTHER, Specify" variables having values of blank (Section 2.2.1) or bad data (i.e., where the bad data code denoted a nonsensical answer that the respondent keyed). These edits were *not* implemented if respondents had "OTHER, Specify" responses of "don't know" or "refused" in addition to specification of OTCs; such responses were interpreted to mean that the respondent was still a potential nonmedical user of some prescription-type medication, especially in situations where respondents may not have known what they ingested.

---

[62] In the pain relievers and stimulants sections, this edit also involved assigning codes of 81 to the lifetime OxyContin® variable OXYCONTN and the methamphetamine variable METHDES, respectively, even though lifetime nonuse was not really logically inferred. In the pain relievers module, respondents already would have answered the lead question PR04 as "no," indicating that they had never used any prescription pain relievers below the red line on Pill Card A, including OxyContin®. Similarly, respondents would already have answered the lifetime methamphetamine question ST01 as "no," indicating that they never used methamphetamine, Desoxyn®, or Methedrine®. However, the code of 81 was assigned to OXYCONTN or METHDES for consistency with the assignment of codes of 81 to the other respective pain reliever or stimulant gate variables.

If other qualifying prescription-type medications were specified in addition to OTCs, the respondent's status as a nonmedical user was retained (e.g., if a respondent reported nonmedical use of a prescription pain reliever in addition to use of aspirin in the pain relievers section of the interview). Further, the OTC responses were retained in the respondent's "OTHER, Specify" variables. If a respondent reported use of a drug that may be available over the counter in certain strengths but is available in other strengths only by prescription, then the respondent's status as a nonmedical user of that category of prescription-type psychotherapeutics did not change. For example, specification of ibuprofen or Motrin® without a dosage could refer to use in prescription form, and this was assumed to be the case in the editing. However, specification of Advil® (i.e., an OTC dosage of ibuprofen) would be an unambiguous indication of use of an OTC drug. Certain drugs were treated as OTCs if they at one time had been available only by prescription but have become available over the counter without a prescription-strength counterpart (e.g., Benadryl®).

6.2.2 Edits of Recency-of-Use Variables

Edits of the variables that establish when respondents last used a drug of interest are probably the most critical.[63] These recency-of-use variables are the precursors for the final measures that establish the prevalence of use in the past 30 days, past 12 months, and lifetime.

The skip logic in the CAI instrument limited the kinds of information that were available for use in editing the recency-of-use variables. In particular, respondents who answered a gate question (or all multiple gate questions) as "no" (i.e., never used that drug) were not given the opportunity to answer additional questions as though they were users of that drug. Similarly, respondents who reported that they last used a drug "more than 12 months ago" were not given the opportunity to answer further questions in that module about use in the past 12 months or past 30 days, as though they were more recent users than what they had originally indicated (Figure 6.1).

6.2.2.1 Edits to Recency Variables for Nonusers

As noted in Sections 2.2.3 and 2.4.2 in Chapter 2, respondents were skipped out of all remaining questions about use of a particular drug when they answered "no" to a given gate question for cigarettes through heroin or when they answered "no" to all of the individual gate questions for hallucinogens through sedatives.

Codes of 91 (i.e., "never used") typically were assigned to the core recency variables when it was determined unambiguously that respondents had never used the drug of interest. For crack cocaine, this situation also held if the lifetime variable CRKEVER was coded as 91 because the respondent reported never using cocaine in any form. Similarly, for hallucinogens through sedatives, the recency variables were assigned codes of 91 if the all of the lifetime use variables corresponding to the gate questions had been assigned codes of 91 because respondents had never used any of the individual drugs (Section 6.2.1.2).

[63] For brevity, the term "use" in the remainder of Section 6.2 also refers to nonmedical use of prescription psychotherapeutic drugs.

However, editing of the OxyContin® recency variable OXYCREC was handled somewhat differently because the lifetime OxyContin® variable OXYCONTN came from an "enter all that apply" item. In turn, respondents' reports of whether they had ever used OxyContin® nonmedically governed whether they were asked the OxyContin® recency question. Therefore, if respondents had never used any prescription pain reliever nonmedically (and OXYCONTN = 91), then OXYCREC was assigned a code of 91 because the data were conclusive that the respondent had never misused any prescription pain relievers, including OxyContin®. Similarly, OXYCREC was assigned a code of 91 if the edited variable ANLCARD was answered as "no" (i.e., ANLCARD = 2). Because a picture of OxyContin® was shown below the red line on Pill Card A, a response of "no" in ANLCARD was taken as an unambiguous indication that the respondent had never used OxyContin® nonmedically.

In comparison, if the lifetime OxyContin® variable OXYCONTN had a code of 6 ("Response not entered"; see Section 2.4.4), then that was not as strong of an indication that the respondent had never used OxyContin® nonmedically, compared with questions in which respondents explicitly were required to answer "yes" or "no" regarding whether they had ever used the drug of interest. Therefore, when OXYCONTN = 6 because OxyContin® was not chosen from the list of drugs in question PR04A, the recency variable OXYCREC was assigned a code of 81 (NEVER USED OXYCONTIN Logically assigned), as opposed to a code of 91. That is, it was *logically* inferred that respondents had never used OxyContin® when respondents reported nonmedical use of some pain relievers from the PR04A list but not OxyContin®. These indications of 81 were treated the same as codes of 91 in other edits. However, use of the code of 81 introduced the additional detail for analysts that a logical inference had been made about respondents having never used OxyContin®. In contrast, if respondents had been asked, "Have you ever, even once, used OxyContin…" and they answered "no," it would not be necessary to make a logical inference because the respondents would have explicitly reported that they never used it nonmedically.

In addition, codes of 81 were assigned to the relevant edited recency variables for pain relievers (ANALREC), tranquilizers (TRANREC), stimulants (STIMREC), or sedatives (SEDREC) if respondents' only reported nonmedical use of prescription drugs in that module involved use of OTCs (Section 6.2.1.5). Codes of 81 also were assigned to the recency variables for OxyContin® (OXYCREC) and methamphetamine (METHREC) if respondents' only reported nonmedical use of pain relievers or stimulants, respectively, involved OTCs. Although respondents would not have been lifetime nonmedical users of OxyContin® or methamphetamine in order to report lifetime nonmedical use of only OTCs, a code of 81 was assigned to these child recency variables (i.e., rather than a code of 91) for consistency with the code that was assigned to the corresponding parent recency variable.

6.2.2.2 Recency Periods in the Questionnaire

In the core modules for alcohol through sedatives, the following standard codes for recency applied, based on the available responses to the question(s) about most recent use:

1 = Within the past 30 days,

2 = More than 30 days ago but within the past 12 months, or

3 = More than 12 months ago.

In the sections of the tobacco module for cigarettes, snuff, chewing tobacco, and cigars, lifetime users first were asked whether they used the relevant tobacco product in the past 30 days (e.g., question CG05 for cigarettes). If respondents reported that they did not use the particular tobacco product in that period, they were asked to report most recent use prior to the past 30 days (e.g., question CG06 for cigarettes). Recency questions for these tobacco products included categories for most recent use (1) more than 30 days ago but within the past 12 months; (2) more than 12 months ago but within the past 3 years; and (3) use more than 3 years ago.[64]

For these tobacco products, single edited recency variables were created that combined the data for the questions about use in the past 30 days and most recent use in later periods, if applicable. Codes of 1 or 2 in the edited recency variables for these tobacco products had the same meaning as for alcohol through sedatives. In addition, standard codes of 3 of 4 in these edited tobacco recency variables had the following meanings based on respondents' answers to the follow-up question (e.g., question CG06 for cigarettes) if they did not report use in the past 30 days:

3 = More than 12 months ago but within the past 3 years, and

4 = More than 3 years ago.

The CAI instrument included follow-up probes for respondents who were lifetime users of a given drug but did not know or refused to report when they last used it.[65] Respondents who initially did not know when they last used a drug were asked to give their "best guess" of when they last used it. Respondents who initially refused to report when they last used a drug were asked to reconsider answering the question. If respondents changed their initial answer of "don't know" or "refused" to report a definite period when they last used the drug of interest, then that information served as the starting point for subsequent editing of the drug's recency variable. The standard recency codes described previously in this section were assigned according to the specific period of most recent use that was reported in these follow-up probes. In the absence of any inconsistencies between the recency-of-use answers in the follow-up probes and other data within a given drug's module, these answers from the follow-up probes were accepted as final and were incorporated within the edited recency variable.

6.2.2.3 Logically Assigned Recency Periods among Lifetime Users

The following code was assigned to an edited recency variable for alcohol through sedatives if respondents reported lifetime use of a drug but continued on follow-up to answer

[64] A recency variable was not created for pipe tobacco because lifetime users of pipe tobacco were asked only if they smoked tobacco in a pipe in the past 30 days.

[65] For cigarettes, snuff, chewing tobacco, and cigars, these follow-up probes were asked if respondents reported that they did not use a given tobacco product in the past 30 days but they answered "don't know" or "refused" to the follow-up question about most recent use more than 30 days ago.

"don't know" or "refused" regarding when they last used it: 9 = Used at some point in the lifetime LOGICALLY ASSIGNED.

These respondents were eligible to be statistically imputed to be users in any period, including in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago. Similarly, for respondents whose recency-of-use questions for LSD, PCP, Ecstasy, OxyContin®, or methamphetamine had been skipped but they were logically inferred to be lifetime users of these drugs based on their responses to "OTHER, Specify" items (Section 6.2.1.4), their edited recency variables at least initially were assigned a code of 9.

In the recency variables for cigarettes, snuff, chewing tobacco, and cigars, the following code was assigned if respondents reported that they did not use the tobacco product in the past 30 days but they had missing data for the period more than 30 days ago when they last used it: 19 = Used more than 30 days ago LOGICALLY ASSIGNED.

These respondents were eligible to be statistically imputed to have most recently used a given tobacco product more than 30 days ago but within the past 12 months, more than 12 months ago but within the past 3 years, or more than 3 years ago.

The following additional codes could be assigned to the edited recency variables when respondents reported lifetime or more recent use:

8 = Used at some point in the past 12 months LOGICALLY ASSIGNED,

11 = Used in the past 30 days LOGICALLY ASSIGNED,

12 = Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED,

14 = Used more than 12 months ago LOGICALLY ASSIGNED,[66]

29 = Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED, and

39 = Used at some point within the past 3 years LOGICALLY ASSIGNED.

Sections 6.2.2.4, 6.2.2.5, and 6.2.2.6 discuss situations in which these codes were assigned based on the editing procedures. In particular, codes of 11 or 12 could be assigned to a parent recency variable (e.g., any hallucinogen) based on data from a child recency variable (e.g., LSD). Recency variables did not require further statistical imputation when codes of 11 or 12 were assigned.

Recency variables that were assigned codes of 8, 9, 14, 19, 29, or 39 underwent further statistical imputation to assign a specific period of most recent use. These codes were used to place constraints during the subsequent imputation process for assigning the period of most recent use. For example, assigning a code of 8 to an edited recency variable constrained the imputation for the final recency to be within the past 30 days or more than 30 days ago but within the past 12 months.

[66] Codes of 14, 29, and 39 applied only to the recency variables for cigarettes, snuff, chewing tobacco, and cigars.

6.2.2.4 Application of "Flag and Impute" Procedures to Editing of Recency-of-Use Variables

The procedures used to edit the CAI recency-of-use variables were referred to as the "flag and impute" procedures (Section 2.4.6). Under these procedures, the limited situations where potential inconsistencies existed between a respondent's answer to a drug's recency question and other data in that module were identified and flagged.

Table 6.2 lists the usual types of inconsistencies that could occur between a drug's recency variable and other variables in that drug's module, and how these inconsistencies were handled through the flag and impute procedures. In the situations described in Table 6.2, these inconsistencies were handled by statistically imputing final values for the affected recency variable and the other variable(s) where the data were inconsistent with the respondent's original answer to the recency question. Most of the codes that were described in Section 6.2.2.3 were assigned to the edited recency variables as a result of these edits. For example, if a respondent reported first use of a marijuana at his or her current age but also reported most recent use more than 12 months ago, both answers logically cannot be true. In this situation, the edited recency variable MJREC was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) and the inconsistent initiation variables were assigned codes for bad data (Sections 2.3 and 2.4.3).

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables

Type of Inconsistency	Edits Implemented
Recency originally indicates use in the past 30 days, but use on 0 days in the past 30 days is confirmed (suggesting less recent use).	**The edited recency variable was assigned a code of 8** (i.e., Used at some point in the past 12 months LOGICALLY ASSIGNED) to indicate that the respondent (R) was at least a user in the past 12 months and potentially a user in the past 30 days (Section 6.2.2.3). The 30-day frequency was set to missing (i.e., bad data).
Recency originally indicates use more than 30 days ago but within the past 12 months, but the 12-month frequency indicates use on more than 335 days in that period (suggesting past month use).	**The edited recency variable was assigned a code of 8**. The 12-month frequency data were set to missing.
Recency does not indicate use in the past 30 days, but the R reports first using the drug (or smoking cigarettes daily) in the same month as the interview took place (suggesting past month use).	**If the recency originally indicated use more than 30 days ago but within the past 12 months, it was assigned a code of 8. If the recency originally indicated use more than 12 months ago (or was missing), it was assigned a code of 9** (i.e., Used at some point in the lifetime LOGICALLY ASSIGNED) to indicate that the R was at least a lifetime user (and potentially a user in the past 12 months or past 30 days). The values in the month of first use (MFU) and year of first use (YFU) that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes. The MFU and YFU variables also were set to bad data if the recency was missing and had been assigned a code of 9 because the initiation data would have suggested use in the past 30 days.

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables (continued)

Type of Inconsistency	Edits Implemented
Recency does not indicate use in the past 30 days, but the R has other data suggesting initiation of use in the past 30 days (e.g., if first use was indicated at the R's current age and the R's last birthday was fewer than 30 days ago, or based on a comparison of the 12-month frequency and the maximum number of days that the R could have used the drug).	**The edited recency variable was assigned a code of 11** (Used in the past 30 days LOGICALLY ASSIGNED) to infer that the R was a past month user. No further editing was done to the variables indicating use in the past 30 days. This edit was an exception to the general "flag and impute" procedures.
Recency does not indicate use in the past 12 months, but the age at first use (or age at first daily use of cigarettes) equals the R's current age (suggesting past year use).	**For alcohol through sedatives, the edited recency variable was assigned a code of 9**. **For tobacco products, if the lead 30-day question** (e.g., CG05 for cigarettes) **indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19** (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). **Otherwise, the relevant tobacco recency variable was assigned a code of 9.** The age at first use that triggered the inconsistency with the recency-of-use answer was set to bad data. If the month of first use and year of first use were answered (i.e., not blank), the values in these variables were overwritten with bad data codes. (Month and year data were considered to be linked with the age-at-first-use data. Therefore, if the age at first use was questionable, then the month and year were considered to be questionable as well.)
Recency does not indicate use in the past 12 months, but the R reported first using in a month and year that falls within 12 months of the interview date (including data for the month and year when the R reported first smoking cigarettes daily).	**For alcohol through sedatives, the edited recency variable was assigned a code of 9**. **For tobacco products, if the lead 30-day question** (e.g., CG05 for cigarettes) **indicated that the R did not use in the past 30 days, then the recency was assigned a code of 19** (Used more than 30 days ago LOGICALLY ASSIGNED) to indicate that the R was a user more than 30 days ago (but was potentially a user in the past 12 months). **Otherwise, the relevant tobacco recency variable was assigned a code of 9.** The values in the month and year of first use that triggered the inconsistency with the recency-of-use answer were overwritten with bad data codes.
For tobacco products, recency indicates use more than 3 years ago, but age at first use (or first daily use of cigarettes) indicates that the first use was within 2 years of the R's current age (suggesting use within the past 3 years). (Note that edits checking for indications of use at the R's current age were given priority over the condition described here because the former response pattern suggests use in the past year.)	**The recency was assigned a code of 14** (Used more than 12 months ago LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 12 months ago (but potentially in the past 3 years). The age at first use that triggered the inconsistency with the recency-of-use answer was set to bad data. If the month of first use and year of first use were answered (i.e., not blank), the values in these variables were overwritten with bad data codes.

Table 6.2 How the Flag and Impute Edit Procedures Handled Usual Inconsistencies Involving the CAI Recency Variables (continued)

Type of Inconsistency	Edits Implemented
For tobacco products, the R did not use in the past 30 days but did not report a specific period of most recent use more than 30 days ago. However, the R reported first use within 1 or 2 years of his or her current age (i.e., but not at the R's current age).	**The recency was assigned a code of 29** (Used more than 30 days ago but within the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point more than 30 days ago but within the past 3 years (but potentially within the past 12 months).
For tobacco products, the R did not indicate whether he or she used in the past 30 days. However, the R reported first use within 1 or 2 years of his or her current age (i.e., but not at the R's current age).	**The recency was assigned a code of 39** (Used at some point in the past 3 years LOGICALLY ASSIGNED) to indicate that the R last used at some point within the past 3 years (but potentially within the past 30 days or past 12 months).
For chewing tobacco and snuff, the brand of chewing tobacco that Rs reported using most often in the past 30 days was really a snuff brand, or vice versa.	**Created a recoded any smokeless tobacco recency (SLTREC)** that incorporated data from the chewing tobacco and snuff recency variables CHEWREC and SNFREC, respectively. Thus, for example, Rs who reported using chewing tobacco in the past 30 days but specified a snuff brand as the brand they used most often in that period would still be considered a past month user of some type of smokeless tobacco product.

CAI = computer-assisted interviewing.

Note: Indications of most recent use include answers from follow-up probes for the recency questions.

Prior to implementing these flag and impute rules, initial edits checked for situations where a respondent's original answer to an age-at-first-use question might have been inconsistent with his or her recency of use, but a revised age at first use was not. For example, if a respondent reported first using a drug at his or her current age, the respondent was asked to verify this age at first use. If the respondent reported that this age at first use was *not* correct but then on follow-up did not know at what age he or she first used, or refused to answer, the edits updated the age at first use to reflect this "don't know" or refusal response. The rationale for this edit was that the respondent indicated that the initial answer was not correct. A final answer of "don't know" or "refused" to an age at first use question would not necessarily be inconsistent with a reported recency of use more than 12 months ago.

6.2.2.5 Most Recent Use of Smokeless Tobacco

Table 6.2 also lists edits that applied to a special situation for chewing tobacco and snuff. When the CAI instrument was first fielded in 1999, considerable cross-reporting of chewing tobacco and snuff brands was observed among users in the past 30 days, suggesting that respondents were not always clear about the differences between these two types of smokeless tobacco. For example, respondents could report using chewing tobacco in the past 30 days but specify a snuff brand as the brand of "chewing tobacco" they used most often in that period. However, this cross-reporting was identifiable only for respondents who reported use in the past 30 days of either smokeless tobacco product but was assumed to be operating for respondents who reported less recent use. For this reason, a recoded smokeless tobacco recency variable SLTREC was created from the respective chewing tobacco and snuff recency variables

(CHEWREC and SNFREC, respectively). Thus, if a respondent reported use of chewing tobacco in the past 30 days but specified use of a snuff brand in the past 30 days, the respondent was still smokeless tobacco user in that period.

In creating the recoded SLTREC, indications of more recent use of chewing tobacco or snuff were given precedence over indications of less recent use. In situations where one recency variable indicated use in a definite period (e.g., more than 30 days ago but within the past 12 months) and the second recency variable indicated use in an indefinite period (e.g., use at some point in the lifetime, which could have included use in the past 30 days, past 12 months, or past 3 years), the final assignment to SLTREC indicated a less definite recency value. The rationale for this procedure was that the respondent was potentially a user in a more recent period. For example, if a respondent indicated use of chewing tobacco more than 30 days ago but within the past 12 months and the flag and impute rules had assigned a code of 9 to the snuff recency to indicate that the respondent last used snuff at some point in his or her lifetime, the recoded SLTREC was assigned a code of 8 (Section 6.2.2.3) to indicate use at some point in the past 12 months. That is, the report of chewing tobacco use in the past 12 months (but not the past 30 days) could be used to narrow down the use of any smokeless tobacco to some point in the past 12 months, but the respondent could still have used in the past 30 days. Similarly, if one of the recency variables had a missing value but the other did not, the SLTREC variable was assigned a code to indicate that there was some uncertainty about when the respondent last used smokeless tobacco. Suppose, for example, that a respondent reported last using chewing tobacco more than 12 months ago but within the past 3 years, but refused to report whether he or she had ever used snuff. In this situation, the SLTREC variable was given a code to indicate that the respondent used smokeless tobacco at some point in the lifetime because the respondent may have used snuff within the past 12 months or past 30 days.

6.2.2.6 Edits to Parent and Child Recency Variables Other than Smokeless Tobacco

Important exceptions to the general flag and impute principles involved situations where inconsistencies existed between parent and child recency variables (any cocaine and crack cocaine; any hallucinogen use and LSD, PCP, or Ecstasy use; any pain reliever use and OxyContin®; any stimulant use and methamphetamine use). These are presented in Table 6.3, along with a description of how the data were edited when specific types of inconsistencies occurred between related recency variables. In these special situations, indications of use of the child drug (e.g., crack cocaine) that were more recent than that indicated for the parent drug category (e.g., cocaine in any form) were used to logically infer more recent use of the parent drug category. For example, not all respondents might make the connection that crack cocaine fits within the broader category of cocaine in general.

Therefore, if a respondent reported last using any cocaine more than 30 days ago and also reported last using crack cocaine in the past 30 days, the edit procedures assigned a code of 11 to the edited cocaine recency variable COCREC (Section 6.2.2.3). This edit indicated logical inference that the respondent had used cocaine in any form in the past 30 days. Overall, however, imputation played a more prominent role than editing in resolving inconsistencies with respect to the most recent use of a drug.

Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables

Recency Reported by Respondent		Edited Recency	
Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)	**Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)**	**Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)**	**Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)**
(1) Indicates use in past month.	Indicates use that is less recent than the past month.	Retains the recency reported by the respondent (R).	Logically infers the R to be a past month user. Assigns a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED).
(2) Indicates use more than 30 days ago but within the past 12 months.	Coded as 8, indicating use at some point in the past 12 months (Section 6.2.2.3).	Retains the recency reported by the R.	Retains the code of 8 to indicate that the R has used at some point in the past 12 months.
	Indicates use more than 12 months ago.	Retains the recency reported by the R.	Logically infers the R to have last used more than 30 days ago but within the past 12 months. Assigns a code of 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).
	Coded as 9, indicating use at some point in the lifetime (Section 6.2.2.3).	Retains the recency reported by the R.	Logically infers the R to be *at least* a past year user. Assigns a code of 8.
(3) Indicates use more than 12 months ago.	Coded as 8, indicating use at some point in the past 12 months.	Retains the recency reported by the R.	Retains the code of 8 to indicate that the R is *at least* a past year user.
	Coded as 9, indicating use at some point in the lifetime.	Retains the recency reported by the R.	Retains the code of 9 to indicate that the R is *at least* a lifetime user.
(4) Coded as 8, indicating use at some point in the past 12 months (see Table 6.1).	Indicates use more than 30 days ago but within the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.	Assigns a code of 8 to indicate use at some point in the past 12 months.
	Coded as 8, indicating use at some point in the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.	Retains the code of 8 to indicate use at some point in the past 12 months.
	Coded as 9, indicating use at some point in the lifetime.	Retains the code of 8 to indicate use at some point in the past 12 months.	Assigns a code of 8 to indicate use at some point in the past 12 months.

Table 6.3 How the Flag and Impute Edit Procedures Handled Inconsistencies between Parent and Child Recency Variables (continued)

Recency Reported by Respondent		Edited Recency	
Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)	**Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)**	**Child Recency (i.e., crack, LSD, PCP, Ecstasy, OxyContin®, methamphetamine)**	**Parent Recency (i.e., any cocaine, any hallucinogen, any pain reliever, any stimulant)**
(5) Coded as 9, indicating use at some point in the lifetime.	Indicates use more than 30 days ago but within the past 12 months.	Retains the code of 9 to indicate use at some point in the lifetime.	Assigns a code of 8 to indicate use at some point in the past 12 months.
	Indicates use more than 12 months ago.	Retains the code of 9 to indicate use at some point in the lifetime.	Assigns a code of 9 to indicate use at some point in the lifetime.
	Coded as 9, indicating use at some point in the lifetime.	Retains the code of 9 to indicate use at some point in the lifetime.	Retains the code of 9 to indicate use at some point in the lifetime.

Note: These edits take place after inconsistencies have been identified between a recency variable and nonrecency variable (e.g., between the recency and the age at first use). For hallucinogens/LSD/PCP/Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine, these edits also take place after the R has revised one or more answers in response to a consistency check. Further, for hallucinogens/LSD/PCP/Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine, any inconsistencies that remain between a given recency variable and other nonrecency variables following inconsistency resolution are transferred back into the recency variables prior to implementation of these edits. For example, if the original answer to the hallucinogen recency disagreed with the age at first use but the revised recency in response to the consistency check did not, then the recency would be updated to reflect the revised value. Prior to implementation of the edits shown in the table, however, if the revised recency still disagreed with the age at first use, then the recency would be edited further to reflect the fact that the previous inconsistency still remained.

In addition, special patterns could remain in the data for cocaine, hallucinogens, or stimulants after most inconsistencies had been addressed and some related data elements were missing. Specifically, respondents could indicate that they first used any cocaine, any hallucinogen, or any stimulant within 12 months of the interview date (e.g., first use at their current age) and indicate that they last used that drug more than 30 days ago but within the past 12 months (e.g., HALLREC = 2 for any hallucinogen use). Logically, then, if respondents who had used a child drug within a given category (e.g., LSD, PCP, or Ecstasy) had missing data on first use for one or more of their child drugs, it followed not only that they had to have *first* used a child drug at some point in the past 12 months, but they also had to have *last* used the child drug at some point within the past 12 months. In this situation, both the parent recency (e.g., HALLREC) and child recency variable(s) (e.g., LSDREC, PCPREC, or ECSREC) were set to values of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) so that they would be imputed consistently. If the parent recency indicated past month use (e.g., HALLREC = 1) when initiation of the parent occurred in the past 12 months but a child recency (e.g., LSDREC) did not indicate use in the past 12 months, only the child recency was set to a value of 8.

Similarly, respondents could indicate that they first used any cocaine, any hallucinogen, any pain reliever, or any stimulant in the same month that they were interviewed and indicate that they last used the drug in the past 30 days but have missing data on first use for the child

drug(s) within a category. In this situation, respondents who had used a child drug within that category (e.g., LSD) also were inferred to have last used that drug in the past 30 days. The edits assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) to the child recency variables (e.g., LSDREC = 11).

6.2.3 Edits Involving Users of Only Child Drugs

In addition to the situations described in Table 6.2, special edits were applied in situations in which respondents were users of only the child drug (for pain relievers and stimulants) or of only one child drug (for hallucinogens). The following specific patterns indicated that respondents were users of only the child drug:[67]

- for hallucinogens, a gate question for only one of the child drugs (i.e., LSD, PCP, or Ecstasy) was answered as "yes" and all other gate questions for hallucinogens were answered as "no";
- for pain relievers, only the response in question PR04A for OxyContin® was chosen and all other gate questions for pain relievers were answered as "no";
- for stimulants, only the gate question for methamphetamine was answered as "yes" and all other gate questions for stimulants were answered as "no"; or
- for hallucinogens, pain relievers, or stimulants, only the gate question for use of any other drug in the category was answered as "yes" (i.e., including the gate questions for the child drugs) but the *only* report in the "OTHER, Specify" data for that module was for use of a child drug.[68]

Respondents were *not* considered to be users of only the child drug if they reported lifetime use of only the child drug but they had responses of "don't know" or "refused" for any of the other gate questions. For example, if a respondent answered the question about lifetime use of LSD as "yes" and had some responses of "don't know" or "refused" for other hallucinogen gate questions and no reports of lifetime use of hallucinogens from the other gate questions, the respondent could have been a lifetime user of any of the specific hallucinogens for which he or she had missing data.

For hallucinogens, users of only one child hallucinogen who answered the gate question for one of the child drugs as "yes" and answered all remaining gate questions as "no" were asked questions about initiation and recency for the overall parent category. The corresponding questions were skipped for the child drug. This also was the pattern in the data if respondents

[67] For brevity, "use of only the child drug" in the remainder of this section also refers to use of only one child drug for hallucinogens.

[68] Respondents were considered to be users of only the child drug if they had codes only for "blank" or "bad data" in their "OTHER, Specify" data (see Sections 2.2.1 and 2.3) in addition to the report for a child drug. However, respondents were considered *not* to be users of only the child drug if they had codes for "don't know" or "refused" in their "OTHER, Specify" data in addition to the report for the child drug; responses of "don't know" or "refused" were considered to be potential indications of use of another drug besides the child drug. This procedure is consistent with the editing procedures described in Section 6.2.1.5 for identifying respondents who reported use of only OTCs in a prescription drug section.

reported lifetime use only of "other" hallucinogens and specified lifetime use of only a single child drug.

For pain relievers and stimulants, users of only OxyContin® or methamphetamine, respectively, who reported nonmedical use in the relevant gate question (e.g., question ST01 for methamphetamine) were asked questions about initiation, recency, and frequency of use in the past 12 months for the *child* drug. The corresponding questions were skipped for the parent drug. In contrast, for respondents who were identified as nonmedical users of only OxyContin® or methamphetamine based on their "OTHER, Specify" data (i.e., they reported use only of "other" pain relievers or stimulants, respectively but specified only the child drug), the questions were asked for initiation, recency, and 12-month frequency of use of the *parent* drug, and the child drug questions were skipped.

6.2.3.1 Edits to Child Hallucinogen Variables for Users of Only One Child Drug

If respondents reported use of only one child hallucinogen, the values from the parent initiation and recency variables were assigned to the corresponding variables that had been skipped for the child drug. For example, if a respondent reported lifetime use of only LSD, the value from the edited age-at-first-use variable for hallucinogens (HALLAGE) and the value from the edited hallucinogen recency variable (HALLREC) were assigned to the corresponding variables for LSD (LSDAGE and LSDREC, respectively). Similarly, any values for the year and month of first use for recent initiates of any hallucinogen (HALYFU and HALMFU, respectively) were assigned to the corresponding year and month variables for LSD. These values also were retained in the edited variables for use of any hallucinogen. Section 6.2.5 provides further details about the content of the questions in the core drug modules for initiation of use.

6.2.3.2 Edits to Pain Reliever and Stimulant Variables for Users of Only the Child Drug

If respondents reported use of only the child drug and this report was not logically inferred from the "OTHER, Specify" data, the values from the child initiation, recency, and 12-month frequency variables were assigned to the corresponding variables that had been skipped for the parent drug. These values also were retained in the edited variables for nonmedical use of OxyContin® or methamphetamine. For example, if all gate questions for pain relievers except PR04 had been answered as "no" and OxyContin® was the only pain reliever that had been chosen from the list in PR04A, the values from the edited variables for OxyContin® (e.g., OXYCREC for most recent use) were assigned to the corresponding variables for pain relievers (e.g., ANALREC). Section 6.2.4 provides further details about the content of the questions in the core drug modules that established the frequency of use in the past 12 months. As noted previously, Section 6.2.5 provides further details about the content of the questions in the core drug modules for initiation of use, including initiation of nonmedical use of pain relievers or stimulants.

If respondents reported use of only the child drug but this report was logically inferred from the "OTHER, Specify" data, the values from the *parent* initiation, recency, and 12-month frequency variables were assigned to the corresponding variables that had been skipped for the

child drug. For example, if all gate questions for pain relievers except PR05 (nonmedical use of any other pain reliever) had been answered as "no" and OxyContin® was the *only* pain reliever that had been specified, the values from the edited variables for pain relievers (e.g., ANALREC for most recent use) were assigned to the corresponding variables for OxyContin® (e.g., OXYCREC).

6.2.4 Edits of Frequency-of-Use Variables

The CAI instrument included questions about the number of days that respondents used different drugs in the past 30 days or past 12 months (or the average number of days per week or days per month that they used in the past 12 months). These are referred to in this section as 30-day and 12-month frequency variables, respectively. Data from these frequency questions can be used to distinguish between occasional and more frequent users of a drug. For example, frequent users of alcohol and illicit drugs may represent a group who are potentially in need of substance abuse treatment or other services for their substance use. Similarly, regular users of tobacco products, such as people who smoked cigarettes every day in the past 30 days, probably represent a group that would have greater difficulty stopping their use of tobacco. In addition, the alcohol section included a question about the number of days that respondents consumed five or more drinks per occasion in the past 30 days, a question which is used to construct measures of binge and heavy alcohol use in that period.[69]

6.2.4.1 Preferred Ways of Reporting 12-Month Frequency

For the 12-month frequency determinations, respondents first were asked how they preferred to report their frequency of use in the past 12 months. Respondents could indicate a preference to report their frequency of use in one of three ways: (1) use on an average number of days per week in the past 12 months, (2) use on an average number of days per month in the past 12 months, and (3) the total number of days they used in the past 12 months. In particular, respondents who used a drug regularly in the past 12 months might find it easier to report their frequency of use in one of the first two ways as opposed to figuring the total number of days they used in that entire period. Conversely, respondents who used on only a few days in the past 12 months might prefer the third reporting method.

Individual variables (subsequently referred to as "source variables") were created for the preferred way of reporting the 12-month frequency and for the associated follow-up questions (i.e., average number of days per week, average number of days per month, or total number of days in the past 12 months). An overall measure of the number of days that respondents used a substance in the past 12 months was created from these individual source variables.

For respondents who chose to report a total number of days that they used a substance , the 12-month frequency was the actual number of days that the respondent reported using the drug in the past 12 months (assuming no inconsistency with the 30-day frequency; these inconsistencies are discussed in Section 6.2.4.4). For respondents who chose to report an average use in days per week or days per month, the overall number of days that they used in the past 12

[69] Binge alcohol use refers to the consumption of five or more drinks on a single occasion on at least 1 day in the past 30 days. Heavy alcohol use refers to the consumption of five or more drinks on a single occasion on 5 or more days in that period.

months was a calculated value. Specifically, answers in terms of the average number of days used per week in the past 12 months were multiplied by 52, and answers in terms of the average number of days used per month in the past 12 months were multiplied by 12 in order to yield a calculated 12-month frequency. Because these latter two response options were averaged responses over the past 12 months, no further adjustments were made to the calculated 12-month frequency value when respondents used the drug more than 30 days ago but within the past 12 months, and they did not initiate use at some point in the past 12 months.

6.2.4.2 Editing of 12-Month and 30-Day Frequency Variables for Nonusers or Less Recent Users

If the lifetime gate question(s) and edited recency-of-use variable indicated that the respondent had never used the drug of interest,[70] then edits at this step assigned a code of 91 to the 30-day frequency variable (where applicable)[71] and a code of 991 to the final 12-month frequency variable (where applicable). For questions on drugs where respondents were asked to report their frequency of use in the past 12 months, codes of 91 (or 991) were assigned to the source variables pertaining to the preferred method of reporting the 12-month frequency (i.e., average number of days per week, average number of days per month, or total number of days used in the past 12 months), the average number of days per week, the average number of days per month, and total number of days used in the past 12 months.[72]

Similarly, if the edited recency of use indicated that the respondent had used the drug but not in the period of interest, edits at this step assigned a code of 93 to the 30-day frequency variable and codes of 93 (or 993) to the 12-month frequency variable and related source variables that were used to create the 12-month frequency.

6.2.4.3 Editing of 12-Month and 30-Day Frequency Variables When the Recency Was Indefinite

If the respondent was potentially a user in the period of interest (i.e., there was some question about when the respondent last used the drug) and the CAI program had skipped the 30-day or 12-month frequency questions, then the skipped variables retained a blank code. For example, if respondents reported lifetime use of a substance but they did not indicate a specific period for most recent use (Section 6.2.2.3), the frequency-of-use questions retained missing values. If the recency was imputed to indicate use in the past 30 days or past 12 months, the imputation procedures also replaced the missing data in these frequency-of-use variables with non-missing values (See Section 6.3).

Also, if a respondent reported last using marijuana more than 12 months ago, the CAI program skipped the questions pertaining to frequency of marijuana use in the past 12 months

---

[70] For hallucinogens, inhalants, and the psychotherapeutics, this meant that the respondent had never used any of the drugs in that category.

[71] For alcohol, this edit also applied to other 30-day variables, including the variable on the number of days in the past 30 days that respondents had five or more drinks in a single occasion.

[72] If a respondent was logically inferred not to have used a drug and the recency variable had been assigned a code of 81, the corresponding edited 30-day or 12-month frequency variables were assigned codes of 81 or 981, where applicable.

and past 30 days. However, reporting first use of marijuana at the respondent's current age would be inconsistent with the reported recency. As discussed in Section 6.2.2.4 and Table 6.2, the flag and impute edit rules assigned a value to the recency variable to indicate that this respondent was a user at some point in his or her lifetime, which could include use in the past 30 days or past 12 months. In this situation, the 12-month and 30-day marijuana frequency variables that had been skipped retained a blank value in case subsequent imputation might assign the respondent to a more recent category.

In addition, as discussed in Section 2.4.3, if respondents refused the lifetime gate question(s) and were skipped out of the 12-month and 30-day frequency questions (where applicable),[73] the edits at this step assigned refusal codes to the skipped frequency questions (i.e., the refusal was propagated). However, if respondents were skipped out of the 12-month and 30-day frequency questions because they answered the lifetime gate question(s) as "don't know," the edits retained codes of "blank" in the frequency variables, for the reasons given in Section 2.4.3.

6.2.4.4 Editing of the Frequency-of-Use Variables Based on Consistency Checks

Modules that contained both 12-month and 30-day frequency variables included consistency checks between these variables. A consistency check was triggered in situations where the number of days that respondents reported using the drug in the past 30 days exceeded the number of days that the respondent used in the past 12 months.

If respondents did not know how many days they used a drug in the past 30 days or refused to give an answer, they have been asked to give their best estimate of the number of days that they used. Respondents could estimate their 30-day frequency by choosing the category most likely to contain the number of days they used the drug: 1 or 2 days, 3 to 5 days, 6 to 9 days, 10 to 19 days, 20 to 29 days, or all 30 days. A consistency check also was triggered if the number of days that respondents reported using a drug in the past 12 months was lower than the minimum value for the number of days that respondents estimated using that drug in the past 30 days. For example, it would be inconsistent for a respondent to report using marijuana on 6 to 9 days in the past 30 days and also to report using it on fewer than 6 days in the past 12 months.

If the respondent revised either the 12-month or 30-day frequency data (or both) to make them consistent (i.e., such that the 12-month frequency was greater than or equal to the 30-day frequency following any updates done by the respondent), data from the consistency checks were taken as final. This included situations in which respondents resolved inconsistencies between their 12-month frequency and the minimum value for their estimated frequency of use in the past 30 days. If the 30-day frequency (or the minimum value for an estimated 30-day frequency) still was greater than the computed 12-month frequency despite a consistency check having been triggered, then the 12-month frequency was assigned a bad data code.

These edits based on data from consistency checks also applied to data that respondents entered in the follow-up questions for the 12-month and 30-day frequencies for hallucinogens. For example, when respondents were asked follow-up questions about their 30-day frequency of use (Section 6.2.4.6), consistency checks existed between the 12-month hallucinogen frequency

[73] For the tobacco variables through heroin, such a situation would occur if respondents initially refused the gate question and then refused again on follow-up.

and answers to the follow-up questions for the hallucinogen 30-day frequency. Thus, if respondents were routed to the follow-up questions for the hallucinogen 30-day frequency and the resulting 30-day frequency continued to be greater than the 12-month frequency after respondents were prompted to resolve the inconsistency, the hallucinogen 12-month frequency was assigned a bad data code.

In addition, a consistency check was triggered in the alcohol module if respondents reported that they had five or more drinks in a single occasion on more days in the past 30 days than they reported for the number of days in which they drank any alcohol in that period. If respondents made their overall 30-day frequency of alcohol use and frequency of consumption of five or more drinks consistent with one another, then these answers were taken as final.

6.2.4.5 Editing of the Frequency-of-Use Variables in Response to Data Patterns Not Involving Recency, Parent/Child Data, or Binge Alcohol Use

Table B.5 in Appendix B lists detailed edits for the 12-month and 30-day frequency variables. The edits that are described pertain to data patterns that do not involve inconsistencies (1) between the frequency data and most recent use, (2) between frequency-of-use data for parent and child drugs, or (3) involving the frequency of binge alcohol use in the past 30 days. The edits in Table B.5 also pertain to any inconsistent reports that did not trigger a consistency check during the interview or situations in which a consistency check was triggered but respondents did not resolve the inconsistency.

For example, if the value for the 12-month frequency fell within the range of a respondent's estimate for the 30-day frequency (e.g., if a respondent reported using on 8 days in the past 12 months and on 6 to 9 days in the past 30 days), maximum and minimum values were created for the estimated 30-day frequency. In this example where a respondent reported use on 8 days in the past 12 months but estimated using the drug on "6 to 9" days in the past 30 days, use on 6 to 8 days (as opposed to 6 to 9 days) in the past 30 days would be consistent with the respondent's answer to the 12-month frequency. Information on the maximum and minimum possible number of days that a respondent could have used a drug in the past 30 days was used subsequently by the statistical imputation team to assign a final value to the 30-day frequency.

6.2.4.6 Editing of the Frequency-of-Use Variables in Response to Parent/Child Data Patterns for Cocaine, Pain Relievers, and Stimulants

Table B.6 in Appendix B lists detailed edits for the 12-month frequency variables for the parent/child pairs of cocaine and crack, pain relievers and OxyContin®, and stimulants and methamphetamine. This includes description of some special edits to the variable for the preferred way of reporting the 12-month frequency when respondents were nonmedical users of only OxyContin® or users of only methamphetamine (see Section 6.2.3). Table B.6 also lists edits for the 30-day frequency variables for cocaine and crack. Because the hallucinogens module did not include questions on the frequency of use in the past 12 months or past 30 days for the child drugs LSD, PCP, and Ecstasy, inconsistencies in parent/child data for the frequency of use did not occur for hallucinogens.

For example, respondents could report use of a child drug on a number of days in the past 12 months that was greater than the number of days they reported using the parent drug in the past 12 months. In these situations, the higher value from the 12-month frequency for the child drug was assigned to the 12-month frequency for the parent drug. The source variables for the parent 12-month frequency also were edited to indicate the movement of data from the 12-month frequency of the child drug to the 12-month frequency for the parent drug.

Also, since 2003, the hallucinogens and stimulants modules have included follow-up questions for the 12-month frequency variables. Similar follow-up questions have been included in the pain relievers module since 2005. In the hallucinogens module, for example, these questions were asked if respondents originally reported that they last used any hallucinogen more than 12 months ago but subsequently reported more recent use of any hallucinogen, LSD, PCP, or Ecstasy. Respondents' original answer of use of the parent more than 12 months ago would cause them to be skipped out of the 12-month frequency-of-use questions. Therefore, when respondents gave some updated indication of use in the past 12 months, they were asked to fill in previously missing information about their frequency of use in the past 12 months. Similar logic was in place in the pain relievers and stimulants modules. For example, respondents were asked follow-up questions for their frequency of use of any pain relievers in the past 12 months if they originally reported that they last used any pain reliever more than 12 months ago but they subsequently indicated that they last used OxyContin® in the past 12 months, and they continued to be past year users when prompted to resolve the inconsistent data for their period of most recent use of any pain relievers and OxyContin®.

In addition to the 12-month frequency follow-up questions, the hallucinogens module had similar follow-up questions for the 30-day frequency of use. These questions were asked when respondents originally indicated that they last used any hallucinogen more than 30 days ago but subsequently reported that they last used any hallucinogen, LSD, PCP, or Ecstasy within the past 30 days. Thus, since 2003, data on the frequency of hallucinogen use in the past 30 days have been intended to be supplied by respondents (instead of through statistical imputation) if they subsequently indicated some hallucinogen use in the past 30 days but did not originally report using hallucinogens in the past 30 days.

6.2.4.7 Editing of the Frequency-of-Use and Related Variables Involving Binge Alcohol Use

Table B.7 in Appendix B includes edits related to the data for the frequency of binge alcohol use in the past 30 days. For example, question AL07 in the alcohol module (edited variable NODR30A) asked respondents to report the usual number of drinks that they consumed in a given day in the past 30 days. Although this is not a frequency variable per se, information from this variable was used to edit the 30-day frequency data for alcohol and the frequency data for binge alcohol use. For example, if a respondent reported having five or more drinks per occasion on *exactly* the same number of days that he or she reported drinking any alcohol in the past 30 days, then it would logically follow that the respondent's usual number of drinks per day had to have been five or more. If the respondent reported usually having fewer than five drinks on the days when he or she drank alcohol in the past 30 days, NODR30A was assigned a special code of 975 (AT LEAST 5 Logically assigned) to indicate usual consumption of at least five drinks.

Similarly, if a respondent drank on only 1 day in the past 30 days and reported having fewer than five drinks on that 1 day in question AL07, but the respondent answered question AL08 as "don't know" or "refused," it would logically follow that the respondent could not have had five drinks on any occasion in the past 30 days. When this occurred, the edited variable DR5DAY (corresponding to question AL08) was assigned a special code of 80 (NO OCCASIONS OF 5 OR MORE DRINKS IN THE PAST 30 DAYS Logically assigned) to denote that the respondent logically could be inferred not to have had five or more drinks on an occasion in the past 30 days. Consequently, the respondent could subsequently be defined as not being a binge alcohol user in the past 30 days.

6.2.4.8 Frequency of Smokeless Tobacco Use

Although a recoded recency variable SLTREC was created for any smokeless tobacco use, a variable was not created for the number of days that respondents used any smokeless tobacco in the past 30 days. If respondents reported use of both chewing tobacco and snuff in the past 30 days, it would have been possible for use of both smokeless tobacco types to have overlapped to varying degrees in the past 30 days. However, this degree of overlap was unknown. To create a recoded 30-day frequency of any smokeless tobacco use, assumptions would have been required (e.g., picking the maximum of the two) that could not have been confirmed from the data.

6.2.5 Incidence (Age at First Use, Month at First Use, and Year at First Use)

In all core modules except for pipe tobacco, respondents were asked how old they were when they first used the drug of interest. If respondents reported first using the drug within 1 year of their current age, they were asked to report the specific month and year when they first used, with the allowable years ranging from 2010 to 2012. If respondents reported first using the drug at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the first use of the drug occurred in the current year (i.e., 2012). These respondents were asked only for the month that they first used in the current year. The remaining respondents who first used a drug within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They then were routed to a question to report on the specific month that they first used the drug in the year they had reported previously.

Because the routing logic to the different versions of the month-of-first-use and year-of-first-use questions was mutually exclusive, a single, composite set of month-of-first-use (MFU) and year-of-first-use (YFU) variables was created from the individual unedited variables. In addition, if respondents indicated a specific year that they first used a drug, the final year-of-first-use variables for 2012 were recoded to replace unedited codes with values for the specific years (i.e., 2010 through 2012). If respondents confirmed that they first used a drug at their current age and were interviewed subsequent to their birthday, a code of "2012" to the year of first use; this was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the month- and year-of-first-use questions had been skipped because respondents first used the drug more than 1 year younger than their current ages, legitimate skip codes were assigned to the final MFU and YFU variables.

Since 2002, consistency checks have been included in the instrument if the values for the MFU and YFU were inconsistent with the age at first use (AFU). Specifically, for recent initiates of a given drug, the CAI program calculated a second AFU use based on the MFU and YFU by comparing these data with the respondent's date of birth. This comparison was not done if the respondent reported first use of the drug in the same month that he or she was born; a unique AFU could not be determined from the MFU and YFU in these situations because it was not known whether the drug use occurred before or after the respondent's birthday. Similarly, a consistency check was not triggered if the respondent had missing data in either of the month or year questions, such as if the respondent knew the year when he or she first used a drug but did not know the MFU.

In remaining situations in which respondents provided complete data for the MFU and YFU, a consistency check was triggered if the MFU and YFU suggested that respondents initiated use of the drug at an earlier or a later age than what they had previously reported. For example, a consistency check was triggered if a 16-year-old respondent reported first using a drug at age 16 but then reported first using the drug in a month and year that would have meant the respondent was 15 years old when he or she first used the drug. No editing needed to be done if respondents indicated twice in a row that the AFU that was calculated from the MFU and YFU was correct. The CAI program updated the value for the AFU (e.g., AGE1STCG for cigarettes) to agree with the values for the MFU and YFU.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their AFU (e.g., question CG04 for cigarettes) was correct, they had an opportunity to revise the values for their YFU and their MFU. If a consistency check was triggered between the AFU and data in the MFU and YFU, the MFU and YFU were updated with any year and month data that the respondent entered in the consistency checks (e.g., CGCC21 and CG221a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the month- and year-of-first-use data were picked up from the original source variables (e.g., CG04A through CG04D for any cigarette use) for use in editing, such as for the "flag and impute" edits that were described in Section 6.2.2.4 and Table 6.2.

Table B.8 in Appendix B lists edits pertaining to the incidence variables when consistency checks had been triggered because of inconsistent data between the AFU and the YFU and MFU. The default when a respondent did not resolve an inconsistency between the AFU and the MFU and YFU was to favor the AFU in subsequent editing decisions. Table B.9 presents information on additional edits involving the incidence variables other than for parent/child relationships, such as situations where the age at first use was inconsistent with the respondent's current age. Table B.10 presents information on edits involving parent/child relationships in the incidence data. In addition to parent/child relationships that have been described previously for the cocaine, pain relievers, and stimulants modules, parent/child relationships for incidence included first use of any cigarette and initiation of daily cigarette use.

Also, recoded variables were created for respondents' ages when they first used any smokeless tobacco product (i.e., chewing tobacco or snuff), and the month and year when they first used, if applicable. If respondents had a missing value for one of the types of smokeless tobacco (i.e., "don't know," "refused," "bad data"), the missing value was retained in the recoded smokeless tobacco variables for the AFU, YFU, or MFU. For example, if a respondent had used

both chewing tobacco and snuff and reported an age when he or she first used chewing tobacco but refused to report the age when he or she first used snuff, the respondent may have used snuff at a younger age than was reported for chewing tobacco. If the AFU did not have a missing value, the recoded MFU and YFU for smokeless tobacco subsequently were edited to be consistent with the AFU that was chosen. If respondents initiated use of both types of smokeless tobacco at the same age and were asked the month and year that they first used (i.e., the first use was within 1 year of their current age), the recoding procedures picked the earliest year. If they reported first using both types of smokeless tobacco in the same year, the recoding procedures picked the earliest month.

6.2.6 Editing of Noncore Methamphetamine Use Data

As noted in Section 6.1.3, questions have been included in the noncore special drugs module since 2005 to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Additional follow-up items have been included since 2006 to identify those respondents who specifically did not report methamphetamine use in the core stimulants module because they did not consider methamphetamine to be a prescription drug. This section describes the editing procedures for the noncore methamphetamine data in the special drugs module that were used to create the CPN variables for methamphetamine and stimulants. Editing of methamphetamine variables from the special drugs module that was not relevant to creating the CPN variables is discussed in Section 7.4.1 in Chapter 7.

Figure 6.2 shows routing logic for the methamphetamine and stimulant variables in the special drugs module that were used to create the edited CPN variables for most recent use of methamphetamine and stimulants. The figure also includes information about decision making for creating the edited CPN recency variables.

Table 6.4 shows the mapping of questions in the noncore special drugs module to edited variables that were used to create the edited CPN recency variables for methamphetamine and stimulants. In particular, questions SD17A through SD18B module captured information about methamphetamine use from respondents who did not report methamphetamine use in the core stimulants module. Data from questions SD10A, SD10B, SD10C, and SD11 that were present in the special drugs module prior to 2005 also were used in creating the edited CPN variables. Unlike the questions that were described previously, SD10A and SD10B were asked if respondents reported lifetime use of methamphetamine in the core stimulants module.

The general issues that were described in Section 6.2 for the core recency variables also applied to the CPN variables. For example, if respondents reported lifetime methamphetamine use in question SD17A (edited variable MTHAMP) but did not know or refused to report in question SD17B when they last used it, the edited methamphetamine recency variable MTHAREC was assigned a code of 9 (Used at some point in the lifetime, LOGICALLY ASSIGNED).

Table 6.4 Mapping of Noncore Methamphetamine Questions in the Special Drugs Module to Edited Variables

Question(s)	Edited Variable	Comments
SD10A	MTHNEEDL	Ever used a needle to inject methamphetamine; asked if respondents (Rs) reported methamphetamine use in the core stimulants module.
SD10B	MTNDLREC	Most recent use of methamphetamine with a needle, if SD10A was answered as "yes."
SD10C	OSTNEEDL	Ever used a needle to inject stimulants/any other stimulant; asked if Rs reported nonmedical use of stimulants in the core stimulants module. SD10C was worded as "any other stimulant" if lifetime methamphetamine use was reported and was worded as "any stimulant" if methamphetamine use had not been reported.
SD11	OSTNLREC	Most recent use of stimulants/other stimulants with a needle, if SD10C was answered as "yes."
SD17A	MTHAMP	Ever used methamphetamine; asked if Rs did not report lifetime use in the core stimulants module.
SD17B	MTHAREC	Most recent use of methamphetamine, if SD17A was answered as "yes."
SD18A	MTHANEDL	Ever used a needle to inject methamphetamine; asked if Rs did not report lifetime use in the core stimulants module.
SD18B	MTANDLRC	Most recent use of a needle to inject methamphetamine, if SD18A was answered as "yes."
SD17A1 SD17ALT	MTHEVCK	Consistency checks to determine which answer was correct: the report of lifetime nonuse of methamphetamine from the core stimulants module or the report of methamphetamine use in the special drugs module.
SD17A2	MTHNORSN	Reason for not previously reporting methamphetamine use in the core stimulants module; asked if Rs confirmed their use of methamphetamine.
SD17A2SP	MTHNOSP	"OTHER, Specify" response for not reporting methamphetamine use in the core stimulants module; asked if Rs reported "some other reason" for not reporting use.

Figure 6.2 Routing Logic and Editing Overview for Methamphetamine and Other Stimulant Questions in the Special Drugs Module

METH = methamphetamine; STIM = stimulants.

6.2.6.1 Editing of the Methamphetamine Recency Variables in the Special Drugs Module

Table B.11 in Appendix B includes edits for the noncore methamphetamine and stimulant variables that were relevant for creating the edited CPN recency variables. The focus is on edits pertaining to most recent use of methamphetamine (SD17B) or most recent use of methamphetamine or stimulants with a needle (SD10B or SD18B for methamphetamine; SD11 for other stimulants). However, Table B.11 also includes details about editing of the needle use variables—including logical inference of use or nonuse of methamphetamine or other stimulants with a needle—to allow all of the issues for editing of these variables to be included in the same table. In particular, if respondents had never used a needle to inject methamphetamine or other stimulants, then the edited variables pertaining to most recent use of methamphetamine or stimulants with a needle did not contribute to the creation of the edited CPN recency variables.

In contrast, it was possible for respondents to report in the "OTHER, Specify" questions SD05A through SD05E pertaining to use of other drugs with a needle that they had injected methamphetamine or other stimulants at some point in their lifetime. These responses overruled any denial of methamphetamine use (or use of methamphetamine with a needle) or any denial of use of stimulants with a needle. In these situations, the general practice was to assign a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), with the following exceptions:

- If respondents were lifetime users of methamphetamine in the core stimulants module but the most recent use of methamphetamine from the variable METHREC was more than 12 months ago, the respondent was logically inferred to have last used a needle to inject methamphetamine more than 12 months ago. MTNDLREC was assigned a code of 13 (More than 12 months ago LOGICALLY ASSIGNED).
- If respondents were routed to question SD17A because they did not report methamphetamine use in the core stimulants module and the most recent use of methamphetamine from MTHAREC (corresponding to SD17B) was more than 12 months ago, then MTANDLRC was assigned a code of 13 to indicate that the respondent logically had last injected methamphetamine more than 12 months ago.
- If respondents were logically inferred to have used a needle to inject other stimulants in their lifetime but the most recent use of stimulants from the variable STIMREC was more than 12 months ago, then OSTNLREC was assigned a code of 13 to indicate that the respondent logically had last used a needle to inject other stimulants more than 12 months ago.

6.2.6.2 Creation of the Edited CPN Recency Variables for Stimulants and Methamphetamine

In creating the CPN recency variables for methamphetamine and stimulants, the core recency variables METHREC and STIMREC were used as the "base" variables. Consequently, if data from the special drugs module did not indicate use or more recent use than was indicated in the core recency variables, then the CPN recency variables retained the information from the core variables.

The following edited CPN recency variables were relevant to determining most recent use of methamphetamine and stimulants if respondents had reported use of methamphetamine in the core stimulants module:

- MTHREC04: Most recent use of methamphetamine, based on the core recency variable METHREC and the variable MTNDLREC for most recent use of methamphetamine with a needle; and
- STMREC04: Most recent use of stimulants, based on the core recency variable STIMREC and the variable OSTNLREC for most recent use of (other) stimulants with a needle.

In these variables, the number "04" represented editing based on variables that were in the special drugs module in 2004 (i.e., prior to addition of the follow-up questions in 2005).

In creating MTHREC04 and STMREC04, MTHREC04 was initially set to the value from METHREC, and STMREC04 was initially set to the value from STIMREC. The following edits were implemented for MTHREC04 and STMREC04:

- If respondents were lifetime nonusers of methamphetamine in the core stimulants module and respondents did not specify in the special drugs module that methamphetamine was "some other drug" that they injected with a needle, then MTHREC04 retained the value of 91 (Never used) from METHREC. A similar edit applied to STMREC04 if respondents had reported never using stimulants nonmedically and did not specify injecting stimulants as "some other drug."
- If METHREC and STIMREC had been assigned codes of 81 (Never used; logically assigned) because respondents reported use of only OTCs in the stimulants module (Section 6.2.1.5), then MTHREC04 and STMREC04 retained codes of 81. However, this edit did not apply if respondents reported in the special module that these were "some other drug" that they injected.
- Respondents were logically inferred to be more recent users in MTHREC04 and STMREC04 if the needle recency variables MTNDLREC and OSTNLREC indicated more recent use of these drugs with a needle than was indicated in the corresponding core recency variables METHREC and STIMREC. Codes of 11 (Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED) were assigned to MTHREC04 and STMREC04 based on these edits.
- If METHREC indicated use more than 12 months ago or nonuse and MTNLDREC had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) because respondents specified use of methamphetamine as "some other drug" that they injected, then MTHREC04 was set to 9. A similar edit was applied to STMREC04. In addition, if MTHREC04 had been set to 9 because of these edits but STIMREC indicated use more than 12 months ago, then STMREC04 was set to 9.
- If METHREC indicated use more than 30 days ago but within the past 12 months and MTNLDREC had been set to a value of 9, then MTHREC04 was set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED). A similar edit

was applied to STMREC04. In addition, if MTHREC04 had been set to 9 because of the edits that were described previously but STIMREC indicated use more than 30 days ago but within the past 12 months, then STMREC04 was set to 8.

- If MTHREC04 indicated more recent use of methamphetamine (including use of methamphetamine with a needle) than was indicated in STIMREC or OSTNLREC, then STMREC04 was assigned a value indicating more recent use.

The following edited CPN recency variables were the final CPN recency variables that were used in the subsequent imputation procedures that are described in Section 6.3.7:

- MTHREC06: Most recent use of methamphetamine, based on MTHREC04, MTHAREC, or MTANDLRC; and
- STMREC06: Most recent use of stimulants, based on STMREC04 and MTHREC06.

In these variables, the number "06" represented editing based on variables that have been included in the special drugs module since 2006. MTHREC06 was initially set to the value from MTHREC04 and STMREC06 was initially set to the value from STMREC04 to capture the results of any editing when respondents had reported lifetime use of methamphetamine in the core stimulants module.

As noted in Section 6.1.3, however, giving respondents a second opportunity to report methamphetamine use could bias the estimates if respondents who had made a mistake in answering the previous question about methamphetamine use in the core stimulants module could change their answer on follow-up to indicate use—something that is not done for other drugs in NSDUH. Rather, the aim of asking the follow-up methamphetamine questions if respondents had not previously reported use in the core stimulants module was to identify respondents who had not reported methamphetamine use in the context of questions about *prescription* stimulants. Therefore, if respondents who previously did not report methamphetamine use confirmed in the special drugs module that they were indeed users, they were asked why they had not reported methamphetamine use when they were asked about it earlier in the stimulants module. Respondents could indicate one of the following reasons why they had not previously reported methamphetamine use:

- the earlier question in the core stimulants module asked about prescription drugs, and they did not think of methamphetamine as a prescription drug;
- they made a mistake when they answered the earlier question about methamphetamine; or
- there was some other reason.

If respondents reported that there was some other reason why they had not previously reported methamphetamine use, they were asked to specify what this other reason was.

In creating MTHREC06, only those respondents who were routed to question SD17A and indicated that they did not think of methamphetamine as a prescription drug (or who specified something to that as their other reason for not previously reporting use) were counted as additional methamphetamine users. Otherwise, data from MTHAREC and MTANLREC were

not used in creating MTHREC06 if respondents reported that they made a mistake in not previously reporting methamphetamine use in the core or the other reason they specified did not pertain to their not thinking of methamphetamine as a prescription drug. However, if respondents had specified using methamphetamine with a needle as "some other drug," then this was reflected in the edits that were mentioned previously for MTHREC04. In turn, these edits to MTHREC04 affected the final value in MTHREC06.

The following edits were implemented for MTHREC06 based on data from questions SD17A to SD17ASP. These edits that are described for MTHREC06 when respondents reported methamphetamine use assume that respondents had not reported methamphetamine use in the core stimulants module because they did not think of it as a prescription drug.

- If respondents were lifetime nonusers of methamphetamine in the core stimulants module, they indicated in SD17A that they never used methamphetamine, and they did not specify use of methamphetamine with a needle, then MTHREC06 retained the value of 91 (Never used) from METHREC.
- If MTHAREC had been set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) because respondents reported in SD17B that they had last used methamphetamine more than 30 days ago but within the past 12 months and MTANDLRC had a value of 9, then MTHREC06 was assigned a value of 8.
- If MTHAREC had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) because respondents reported in question SD17B that they last used methamphetamine more than 12 months ago but MTANDLRC had a value of 9, then MTHREC06 was assigned a value of 9.
- If the methamphetamine recency variables were consistent (i.e., MTHAREC indicated as recent or more recent use of methamphetamine than was indicated in MTANDLRC, or respondents never used a needle to inject methamphetamine), then the value from MTHAREC was assigned to MTHREC06.
- If MTHAREC indicated that respondents were logically inferred to be more recent users of methamphetamine based on indications of more recent use of methamphetamine with a needle in MTANDLRC (Table B.11), then MTHREC06 was assigned the corresponding code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED) from MTHAREC.

The following edits were implemented for STMREC06 based on data from MTHREC06 according to the edits that were described previously:

- If MTHREC06 had been set to a value of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) and STMREC04 did not indicate use in the past 30 days, then STMREC06 was assigned a value of 8.
- If MTHREC06 had been set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) and STMREC04 did not indicate use in the past 12 months, then STMREC06 was assigned a value of 9.

- If MTHREC06 was consistent with STMREC04 (i.e., STMREC04 indicated as recent or more recent nonmedical use of stimulants than was indicated in MTHREC06, or MTHREC06 indicated that respondents never used methamphetamine), then the value from STMREC04 was assigned to STMREC06.
- If MTHREC06 and STMREC04 indicated that they last used methamphetamine or stimulants in a definite period (i.e., in the past 30 days, more than 30 days ago but within the past 12 months, or more than 12 months ago) and MTHREC06 indicated more recent use of methamphetamine, then STMREC06 was assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED) or 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED).

If MTHREC06 indicated that respondents last used methamphetamine in a definite period but STMREC04 indicated nonuse or had a missing value, then STMREC06 was assigned a code of 11, 12, or 13 (Used more than 12 months ago LOGICALLY ASSIGNED) based on the most recent use from MTHREC06. This included situations in which STMREC04 retained a code of 81 81 (Never used; logically assigned) from STIMREC because respondents reported use of only OTCs in the stimulants module (Section 6.2.1.5). In this situation, the data from MTHREC06 (based on MTHAREC) indicating methamphetamine use overruled the logical inference that respondents were never lifetime nonmedical users of stimulants.

6.2.7 Editing of Lifetime Daily Cigarette Use Status

Lifetime users of cigarettes were asked in question CG15 whether they ever had a period when they smoked cigarettes every day for at least 30 days. Respondents were not asked question CG15 if they had already reported that they smoked cigarettes on all 30 days in the past 30 days.

Edits associated with determining whether respondents had ever smoked cigarettes daily for at least 30 days (edited variable CIGDLYMO) are presented in Table B.12 in Appendix B. In particular, if question CG15 had been skipped because respondents smoked cigarettes on all 30 days in the past 30 days, then CIGDLYMO was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). A code of 5 was assigned rather than a legitimate skip code (Section 2.4.2) to indicate that it could be logically inferred from the skip logic that the respondent had a lifetime period of having smoked cigarettes daily for at least 30 days.

6.2.8 Additional Edits Applied During Imputation Processing

In addition to the logical edits applied as described previously, edits to selected variables were applied during the imputation process and are discussed in this section. In general, these edits affected only a few records. They were implemented mostly to resolve residual inconsistencies that prevented the determination of a valid interval for the assignment of the date of first use (see Section 6.3.3.4).

6.2.8.1 "Other" Hallucinogens, "Other" Pain Relievers, and "Other" Stimulants Variables

For respondents who were known to have never used "other" hallucinogens, "other" pain relievers, and "other" stimulants, certain logical deductions could be made regarding the relationship between the parent drug data and the child drug data if all the necessary conditions that also are described were met.[74] Note that these edits also could have been applied to respondents who were imputed to lifetime nonuse of the "other" variable.

1. If the respondent was known never to have used "other" hallucinogens, the overall hallucinogens recency was missing, and none of the recencies for the child drugs associated with hallucinogens were missing, then the overall hallucinogens recency was assigned to the most recent of its child drug recencies. This also was applied for pain relievers and stimulants.
2. If the respondent was known never to have used "other" hallucinogens, the overall hallucinogens recency was past month, one of the child recencies was past year (where past month vs. not past month use could not be determined), and no other child recency was past month, then the child recency that was past year (where past month vs. not past month use could not be determined) was edited to past month.
3. If the respondent was known never to have used "other" hallucinogens, the parent age-at-first-use value was nonmissing, only one child age-at-first-use value was missing, and the minimum of the nonmissing child age-at-first-use values was greater than the parent age-at-first-use value, then the missing child age-at-first-use value was edited to the parent age-at-first-use value.
4. If the respondent was known never to have used "other" hallucinogens, the parent age-at-first-use value was nonmissing, only one child age-at-first-use value was missing, the minimum of the nonmissing child age-at-first-use values was equal to the parent age-at-first-use value, and the earliest of the nonmissing child months and years of first use was later than the parent month and year of first use, then the missing child age-at-first-use value was edited to the parent age-at-first-use value.[75]

6.2.8.2 Respondents Imputed to Lifetime Use for Child Drug Variables

As discussed in Section 6.3, the first imputation set consisted of the lifetime drug use measures. The results of these imputations could restrict the range of plausible values for other drug use measures, and, therefore, based on this additional information obtained from the imputations, certain editing rules that were applied to the unedited recency and frequency data had to be reapplied. The list of these edits follows:

1. If the parent drug recency of use was known to be lifetime but not past year, and the respondent was imputed to lifetime use of the child drug(s), then the child drug

[74] The creation of these "other" indicators is described in Section 6.1.2.

[75] These cases occur rarely, so they are handled on a case-by-case basis. The procedures do not automatically apply this edit. They flag cases like these for further examination.

recency was set to lifetime but not past year. This was done because the respondent could not have used the child drug more recently than the parent drug.

2. If the respondent used the parent drug on exactly 1 day in the past 12 months, and the respondent was imputed to lifetime use of the child drug, then the child drug recency of use was set equal to the parent drug recency of use, and the child drug 12-month frequency of use was set to 1 day. This was done because the respondent could not have used the child drug on any days when the parent drug was not used, so the recencies and frequencies cannot differ. This edit only applied to OxyContin®, methamphetamine, and crack, which are the only child drugs with 12-month frequencies.
3. If the parent drug incidence data indicated a date of first use in the past year, the parent drug recency of use was past year but not past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year but not past month. This was done because the respondent could not have used the child drug more recently than the parent drug (eliminating the possibility of past month recency), and the respondent also could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency).
4. Similarly, if the parent drug incidence data indicated a date of first use in the past year, the parent drug recency of use was past month, and the respondent was imputed to lifetime use of the child drug(s), then the recency of use for the child drug was set to past year (whether the respondent had used in the past month could not be determined). This was done because the respondent could not have started using the child drug before the parent drug (eliminating the possibility of lifetime but not past year recency).

6.2.8.3 Age-at-First-Use Variables

The edits that are described in this section were implemented in preparation for the imputation process. These edits are designed to reduce processing time and to set the imputation indicators properly (i.e., to "logically assigned" instead of to "statistically imputed"). As with the other edits that were described previously in Section 6.2.8, these edits affected only a small number of records.

1. If the parent age-at-first-use value was missing and the minimum of the child age-at-first-use values was 3 years, then the parent age-at-first-use value was edited to 3 years. This was done because respondents with age-at-first-use values of less than 3 years were ineligible to be donors (see Section 6.3.3.1).[76] This edit applied to all parent age-at-first-use variables: cigarettes, overall hallucinogens, overall pain relievers, overall stimulants, and cocaine.
2. If the parent age at first use was equal to the respondent's current age, all missing child age-at-first-use values were edited to the same age. This edit applied to all child

[76] This could be considered to be an imputation rule rather than an editing rule. Nevertheless, once the decision is made to prevent those with age-at-first-use values of less than 3 years from being donors, it is clear that cases like these do not require any sort of stochastic imputation.

age-at-first-use variables: daily cigarettes, LSD, PCP, Ecstasy, OxyContin®, methamphetamine, and crack.

3. If the parent age at first use was equal to 1 less than the respondent's current age, the child recency[77] was lifetime but not past year (or, for cigarettes, past 3 years but not past year), and the child age-at-first-use value was missing, then the child age-at-first-use value was assigned to 1 less than the respondent's current age. This was done because the child age at first use cannot be less than AGE – 1, because the parent age at first use is AGE – 1, and the respondent could not have begun using a child drug before using the parent drug. It also cannot be greater than AGE – 1, because the child drug recency of lifetime but not past year indicates that the respondent did not use the drug while at his or her current age (because he or she did not use the drug at all in the past year). If the respondent did not use the drug at all in the past year, then he or she could not have begun using the drug in the past year. Because the child age at first use cannot be less than AGE – 1 or greater than AGE – 1, it must be equal to AGE – 1. This edit also applied to all child age-at-first-use variables.
4. If the age at first cigarette use was equal to AGE – 3, cigarette recency was lifetime but not past 3 years, and age at first daily cigarette use was missing, then age at first daily cigarette use was assigned to AGE – 3. The logic is similar to the above edit: the age at first cigarette use precludes the possibility that the age at first daily cigarette use was less than AGE – 3, and the cigarette recency precludes the possibility that the age at first daily cigarette use was greater than AGE – 3.

6.2.9 Edits for Drug Variables that Do Not Undergo Imputation

The following tobacco variables were edited but did not undergo further imputation:

- for adolescents aged 12 to 17 who had never smoked a cigarette: their likelihood of smoking a cigarette if their friends offered them one or of smoking a cigarette in the next 12 months;
- lifetime smoking of 100 or more cigarettes;
- usual brands of cigarettes, snuff, chewing tobacco, or cigars that respondents used in the past 30 days;
- among respondents who smoked cigarettes in the past 30 days:
  - the average number of cigarettes that they smoked per day;
  - the type of cigarette they smoked (light, ultralight, medium, or full flavor);
  - whether the brand of cigarettes that was smoked most often was a menthol cigarette;

[77] Because there was no recency question associated with daily cigarettes, the overall cigarette recency was used instead.

- for respondents who usually smoked Marlboro cigarettes in the past 30 days, the length of the Marlboro cigarettes that they smoked (shorts, regulars or king-sized, or 100s); and
- whether respondents smoked a "roll-your-own" cigarette in the past 30 days.

6.2.9.1 Editing of Likelihood Variables for Adolescent Nonsmokers

If respondents were aged 12 to 17 and had never smoked a cigarette, they were routed to questions CG02 and CG03. Question CG02 (edited variable CIGOFRSM) asked adolescents if they would smoke a cigarette if one of their friends offered them a cigarette. Question CG03 (edited variable CIGWILYR) asked adolescents how likely they thought they would be to smoke a cigarette in the next 12 months.

No editing was done to CIGOFRSM and CIGWILYR if adolescents were routed to the corresponding questions because they never smoked cigarettes. Otherwise, legitimate skip codes of 99 (Section 2.4.2) were assigned if the questions had been skipped because (1) respondents were aged 18 or older, or (2) they were aged 12 to 17 but they had already smoked a cigarette in their lifetime.

6.2.9.2 Editing of Lifetime Smoking of 100 or More Cigarettes

Question CG16A (edited variable CIG100LF) asked cigarette smokers whether they had ever smoked 100 or more cigarettes in their lifetime. Minimal editing was done to CIG100LF. As indicated in Table B.12, respondents were not asked this question if their answers to previous questions indicated that they had smoked 100 or more cigarettes. For example, respondents who smoked cigarettes in the past 30 days were asked in question CG07 to report the number of days that they smoked cigarettes in the past 30 days. Respondents who smoked cigarettes on more than 1 day in the past 30 days were asked in question CG08 to report the usual number of cigarettes that they smoked on those days when they smoked cigarettes. Question CG08 was a categorical variable that gave ranges of numbers of cigarettes that were smoked per day and, where relevant, the equivalent number of packs of cigarettes (e.g., 16 to 25 cigarettes per day, or about 1 pack).

If the product of the number of days that respondents smoked cigarettes and the lower bound of the range was 100 or greater, then question CG16A was skipped. For example, if a respondent smoked cigarettes on all 30 days in the past 30 days and usually smoked about a pack a day (i.e., 16 to 25 cigarettes per day), then a conservative estimate of the number of cigarettes that he or she smoked in the past 30 days (i.e., based on the lower bound) would be $30 \times 16 = 480$.

In this situation, the respondent would not be asked question CG16A because the number of cigarettes that the respondent logically smoked in the past 30 days was greater than 100. As indicated in Table B.12, if question CG16A had been skipped because had been skipped because respondents the quantity and frequency of cigarette use in the past 30 days indicated that respondents had smoked 100 or more cigarettes, then CIG100LF was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).

6.2.9.3 Editing of Tobacco Brand Variables

As noted in Section 2.3.1.2, the CAI instrument included questions to identify the specific brands of tobacco that were used most commonly by respondents who reported use in the past 30 days of cigarettes, chewing tobacco, snuff, or cigars. Respondents could choose from a list of brands or they could indicate use of "a brand not on this list." Procedures for coding the "OTHER, Specify" data when respondents reported use of a brand not on the list were described in Section 2.3.1.2.

The coding approach that was described in Section 2.3.1.2 also applied to situations where respondents chose tobacco brands from the lists of brands that they were shown in the questions. For cigarettes, for example, codes of 101 through 126 were used for Basic through Winston, corresponding to the brands that had been listed in question CG11 prior to the 2005 survey.[78] Cigarette brand codes of 127 through 160 were used for Alpine through True, corresponding to the brands listed in question CG11A prior to the 2005 survey.[79] Thus, the edited variable CIG30BRN for the brand of cigarettes that respondents smoked most often in the past 30 days could be assigned a code of 101 if respondents reported in question CG11 that they usually smoked Basic cigarettes or if they specified Basic cigarettes as "a brand not on this list."

If respondents entered a brand from an available listing, they were asked to confirm their answer. If they confirmed their answer, they were asked no further questions about the brand they used for that particular type of tobacco. However, if respondents indicated that their previous answer was not correct, they were routed back through the series of 30-day brand questions for that type of tobacco. Thus, respondents had the opportunity to make corrections in situations where they may have miskeyed a number, such as if they keyed the number immediately above or below the number of the brand they meant to choose. For each type of tobacco that respondents reported using in the past 30 days, they were allowed to make corrections up to a total of three times. Respondents exited the loop once they confirmed an answer or specified use of a brand not on the list. Respondents also exited the loop if they answered "don't know" or "refused" when asked to confirm their answer.

Because of this routing logic, the brands that respondents confirmed that they used most often in the past 30 days were assigned to the edited variables CIG30BRN (for cigarettes), SNF30BRN (for snuff), CHW30BRN (for chewing tobacco), and CGR30BRN (for cigars). If respondents were rerouted through the series of questions and confirmed their answer on their second or third pass through the questions, the final tobacco brand coding procedures retained the final answer that respondents confirmed and disregarded whatever previous answers the respondent had given but did not confirm. Respondents who answered "don't know" or "refused" when asked to confirm what brand they used were assigned that corresponding code to the final

---

[78] Reference is made to response choices in 2004 or earlier because new response choices were added in 2005. Prior to 2005, for example, the first category in question CG11 was Basic. American Spirit was added to question CG11 in 2005, and this became category 1. To preserve continuity with the codes prior to 2005, these codes were not renumbered in 2005 to reflect any changes to the tobacco brand questions.

[79] The numbering of codes corresponding to responses in question CG11A started with 127 instead of 128 because prior to 2005, the response option of 27 in question CG11 meant "a brand not on this list." This response option was simply a toggle to question CG11A. Because the coding of brands resumed at 127 for brands that were listed in question CG11A, there was no break in the codes.

brand variable for that type of tobacco. If respondents did not confirm what brand of a given tobacco type they used most often in the past 30 days despite three passes through the series of questions, a final code of 9000 to the edited variable for that tobacco type's brand. The code of 9000 indicated that these respondents did not confirm their brand despite three opportunities to do so.

As noted in Section 2.3.1.2, respondents sometimes specified that the brand of chewing tobacco they used most often was actually a snuff brand, or vice versa. Respondents also could specify that the cigarette brand that they smoked most often was actually a brand of little cigars. No editing was done to the codes for tobacco brands that applied to a different type of tobacco. For example, if respondents specified a little cigar brand as the brand of cigarettes that they smoked most often in the past 30 days, the edited variable CIG30BRN retained a code in the 400 or 4000 series for cigars. Continuing this example, if respondents did not report use of cigars in the past 30 days, no editing was done to the data for cigars to indicate use of cigars in the past 30 days in the edited recency variable CIGARREC or to assign the cigar brand code from cigarettes to CGR30BRN.

In addition, a recoded smokeless tobacco brand variable SLT30BRN was created because confusion sometimes existed in terms of what constituted chewing tobacco and snuff. If respondents reported use of both snuff and chewing tobacco in the past 30 days, they were asked to indicate which they had used most often. A final code was assigned to SLT30BRN according to the answer to this question. For example, if a respondent reported that the brand in the chewing tobacco section was the brand that he or she used most often, but this brand of "chewing tobacco" was really a snuff brand, SLT30BRN indicated that the respondent used a particular snuff brand most often in the past 30 days, even though this response came from the chewing tobacco brand questions.

In addition, the precoded response options in the questions for cigarette, chewing tobacco, and cigar brands since 2005 have included more prevalent brands that were mentioned in "OTHER, Specify" data from prior years. Since 2005, response options also are no longer present for less prevalent brands from 1999 to 2004. Documentation of these changes since 2005 is provided in the report on general principles and procedures for editing drug use data in the 2011 MRB (Kroutil, Handley, & Bradshaw, 2013).

These changes in 2005 did not affect the creation of the edited tobacco brand variables. Nevertheless, analysts are advised that this change could affect analyses comparing trends for certain brands prior to 2005 and from 2005 onward. For example, significant differences in prevalence between some tobacco brand estimates prior to 2005 and after 2005 could occur if respondents prior to 2005 needed to type in the brand as "a brand not on this list" but they could choose it from a list in 2005 and beyond.

6.2.9.4 Editing of Miscellaneous Cigarette Use Variables for the Past 30 Days

As noted previously, respondents who reported that they smoked cigarettes in the 30 days prior to the interview were asked a series of additional questions about their cigarette use in that period (i.e., other than the brand of cigarettes that they smoked most often). These questions covered the following topics:

- the average number of cigarettes that they smoked per day (Section 6.2.9.4.1);
- the type of cigarette they smoked (light, ultralight, medium, or full flavor) and whether the brand of cigarettes that respondents smoked most often was a menthol cigarette (Section 6.2.9.4.2);
- for respondents who usually smoked Marlboro cigarettes in the past 30 days, the length of the Marlboro cigarettes that they smoked (shorts, regulars or king-sized, or 100s) (Section 6.2.9.4.3); and
- whether respondents smoked a "roll-your-own" cigarette in the past 30 days (Section 6.2.9.4.4).

6.2.9.4.1 Average Number of Cigarettes Smoked Per Day

Respondents who smoked cigarettes in the past 30 days were asked one of two possible questions regarding the number of cigarettes they smoked per day. Respondents who smoked on only 1 day in the past 30 days were asked to report the number of cigarettes they smoked on that 1 day. Respondents who smoked on more than 1 day (or who estimated the number of days they smoked in the past 30 days) were asked to report the average number of cigarettes they smoked per day. A single, composite variable (CIG30AV) was created from these two questions using data from whatever question the respondents were asked. No further editing was done to the data from these two questions.

6.2.9.4.2 Type of Cigarettes Including Menthol Cigarettes

As for the questions about the number of cigarettes that were smoked per day, past month cigarette users were routed to one of two possible questions about the type of cigarette they usually smoked in the past 30 days (CGTAR1 or CGTAR2). Since 2005, these questions have included a new response option for respondents who smoked "mediums" most often in the past 30 days. Review of data from quarter 1 of 2005 indicated that the addition of this new level for mediums affected the distribution of responses in CGTAR1 and CGTAR2 relative to the distribution in 2004, such that the data would not be comparable between 2004 and 2005. Therefore, the name of the edited variable corresponding to CGTAR1 and CGTAR2 has been CIG30TPE since 2005; prior to 2005, this variable was called CIG30TYP.

Past month cigarette users also were routed to one of two possible questions regarding whether the brand of cigarette they usually smoked was menthol (CGMENTH1 or CGMENTH2). Routing to CGTAR1 or CGTAR2 and to CGMENTH1 or CGMENTH2 was mutually exclusive for respondents who smoked cigarettes in the past 30 days (i.e., respondents were routed to one or the other question in a set but not both). Therefore, composite variables were created for the cigarette type (CIG30TPE) and whether the cigarette brand that respondents smoked most often was menthol (CIG30MEN).

No attempt was made to edit CIG30TPE or CIG30MEN for consistency with the cigarette brand from the variable CIG30BRN. In developing these items, the instrument development team consulted with tobacco research experts regarding which brands offered or did not offer menthol, light, or ultralight varieties. No conclusive information was obtained. As was discussed in Section 2.3.1.2, the "OTHER, Specify" data for tobacco brands did not capture

information for particulars such as regular or menthol forms, or light, ultralight, or full-flavor varieties. Therefore, in situations where respondents specified that level of detail regarding the brand of cigarette that they smoked most often in the past 30 days, that information was not used to edit CIG30TPE or CIG30MEN.

6.2.9.4.3 Type of Marlboro Cigarettes

Respondents who reported in questions CG11, RCG11, or RRCG11 that they usually smoked Marlboro cigarettes in the past 30 days (and who confirmed this report of smoking Marlboro cigarettes) have been asked in question CGLNTH since 2005 about the length of the Marlboro cigarettes they smoked most often in the past 30 days: (1) shorts, (2) regulars or king-sized, or (3) 100s. The edited variable corresponding to CGLNTH was CIG30MLN. CIG30MLN was assigned a legitimate skip code of 99 (see Section 2.4.2) if the cigarette brand variable CIG30BRN did not have a missing value and did not indicate that respondents smoked Marlboro cigarettes most often in the past 30 days; this included situations in which respondents reported that they smoked "a brand not on this list" and did not specify that they smoked Marlboro cigarettes.[80] Because respondents who reported that they smoked some other brand of cigarettes in the past 30 days were not asked CGLNTH, CIG30MLN retained values of 98 (blank) for any respondents who specified that the "other" brand of cigarettes was Marlboro; however, CIG30MLN was blank for this reason for fewer than 25 of approximately 6,500 respondents in 2012 who reported that they smoked Marlboro cigarettes most often in the past 30 days. In addition, CIG30MLN had missing values if respondents who smoked cigarettes in the past 30 days did not know or refused to report at the outset what cigarette brand they smoked most often in the past 30 days, or if they failed to confirm the brand that they smoked most often.

6.2.9.4.4 "Roll-Your-Own" Cigarettes

The cigarette section also included a question (CG14) about whether respondents smoked part or all of a "roll-your-own" cigarette in the past 30 days. The edited variable CIG30ROL corresponded to this question. The cigarette brand question CG11A[81] included response categories for two roll-your-own brands of cigarette tobacco. Respondents who chose either of these roll-your-own brands were skipped out of question CG14; by choosing a roll-your-own brand from the list of cigarette brands, these respondents already had indicated that they had smoked a roll-your-own cigarette in the past 30 days. Therefore, if question CG14 had been skipped and the cigarette brand was one of the roll-your-own brands from CG11a, a code of 5 (Yes LOGICALLY ASSIGNED [from skip pattern]) was assigned to the edited variable CIG30ROL.

---

[80] This assignment of legitimate skip codes included situations in which the "other" cigarette brand was for a different type of tobacco (e.g., if respondents specified that the brand of "cigarettes" they smoked most often was a cigarillo or little cigar) or if respondents reported smoking a brand of cigarettes "not on this list" and they had missing values in their "OTHER, Specify" data. In this latter situation, CIG30BRN continued to have a code of 1999 (Cigarette; brand otherwise unspecified), and it was inferred in the editing of CIG30MLN that this otherwise unspecified cigarette brand was not Marlboro.

[81] Respondents could be routed to questions RCG11, RRCG11, RCG11A, RRCG11A, RCG12, or RRCG12 if they cycled through the cigarette brand questions more than once. For brevity, however, we limit our reference here to the first set of cigarette brand questions: CG11, CG11A, or CG12.

However, respondents could specify a cigarette brand that was not on the list in questions CG11 and CG11A and then specify a roll-your-own brand in question CG12. In this situation, respondents were routed to question CG14. If respondents specified a roll-your-own brand and CG14 was already answered as "yes," no further editing needed to be done. If CG14 was not answered as "yes" when a respondent had specified a roll-your-own cigarette brand, the respondent's original answer was replaced with a code of 3 (Yes LOGICALLY ASSIGNED). This code of 3 signified to analysts that the respondent's original answer in question CG14 had been overwritten to make the roll-your-own data consistent with the cigarette brand information that was recorded in the variable CIG30BRN. However, this edit was implemented in 2012 for fewer than 25 respondents out of more than 15,000 who reported cigarette use in the past 30 days.

6.3 Imputation of the Core Drug Use Variables

The predictive mean neighborhood (PMN) imputation methodology used in the imputation of drug variables beginning in 1999 was applied in a similar manner to the 2012 National Survey on Drug Use and Health (NSDUH) drug data. Consistent with prior years, the drug use measures collected in the 2012 NSDUH included lifetime usage, recency of use, frequency of use in the past 12 months, frequency of use in the past month, and age, year, and month of first use. However, depending on the drug in question, only a subset of these measures were collected and imputed.

Table 6.5 summarizes the drugs and drug use measures that were imputed. This table also indicates how these measures were segregated into units referred to as imputation sets. See Section 3.4 for more information on imputation sets.

Table 6.5 Drugs and Drug Use Measures, Imputation Sets

	Drug Use Measure					
Drug	**Lifetime Usage**	**Recency of Use**	**12-Month Frequency of Use**	**30-Day Frequency of Use**	**Age at First Use**	**Age at First Daily Use**
Cigarettes	NM	Set 2 (12-Month Frequency N/A)			Set 3	Sets 4, 5*
Smokeless Tobacco[1]	Set 1	Set 6 (12-Month Frequency N/A)			Set 7	N/A
Cigars		Set 8 (12-Month Frequency N/A)			Set 9	
Pipes		Set 10 (12-Month and 30-Day Frequency N/A)			N/A	
Alcohol		Set 11[5]			Set 12	
Inhalants		Set 13			Set 14	
Marijuana		Set 15			Set 16	
Hallucinogens[2]		Set 17			Set 18	
Pain Relievers[3]		Set 19 (30-Day Frequency N/A)			Set 20	
Tranquilizers		Set 21 (30-Day Frequency N/A)			Set 22	
Stimulants[4]		Set 23 (30-Day Frequency N/A)			Set 24	
Sedatives		Set 25 (30-Day Frequency N/A)			Set 26	
Cocaine and Crack		Set 27			Set 28	
Heroin		Set 29			Set 30	

Table 6.5 Drugs and Drug Use Measures, Imputation Sets (continued)

Drug	Drug Use Measure					
	Lifetime Usage	Recency of Use	12-Month Frequency of Use	30-Day Frequency of Use	Age at First Use	Age at First Daily Use
Core-Plus-Noncore Stimulants and Methamphetamine	Set 31	Set 32 (12-Month and 30-Day Frequency N/A)			N/A	

N/A = not applicable; NM = never missing. Lifetime cigarette use is used to define a unit respondent and is therefore never missing.

*Prior to imputing age at first daily cigarette use, lifetime daily cigarette use must first be imputed.

[1] Includes chewing tobacco and snuff.
[2] Includes LSD, PCP, and Ecstasy.
[3] Includes OxyContin®.
[4] Includes methamphetamine.
[5] Includes binge drinking frequency.

Because there are numerous sets, some of the set-specific descriptions are explained as deviations from the procedures applied to an earlier set. For example, imputation set 2 is described in detail in Section 6.3.2, and the rest of the recency and frequency sets are described in Section 6.3.5 as deviations from the procedures described in Section 6.3.2.

Because drug use was highly correlated with age, and to facilitate more timely implementation of the imputation procedures, the model building and final assignment of imputed values for all drug use variables were performed separately within three distinct age groups: 12 to 17, 18 to 25, and 26 or older.

6.3.1 Lifetime Drug Use (Imputation Set 1)

The lifetime drug use variables were imputed using the single response propensity (RP)/multiple prediction (PRD) type of PMN, as outlined in Section 3.4.3. In general, the response rates for lifetime drug use variables were very high with less than 1 percent of cases requiring imputation. These high response rates were observed, in part, because of the usable case rule that requires that a respondent answer "yes" or "no" to the question on lifetime use of cigarettes and "yes" or "no" to at least nine additional lifetime use questions.

Because the single RP/multiple PRD type of PMN was used for the lifetime usage imputations, decisions had to be made on the order in which to fit the PRD models. Drugs later in the sequence would have more covariates in their models, because drugs earlier in the sequence were used as covariates after provisional imputation. The order in which the lifetime indicators of use were imputed is shown in Table 6.5, with the exception of lifetime cigarette use.[82] The lifetime use or nonuse of cigarettes was used to define a unit respondent for the NSDUH and, therefore, did not contain any missing values.

[82] See Section 3.4.2 for a brief discussion of how order is determined for imputation sets that use the multiple RP/multiple PRD or single RP/multiple PRD type of PMN.

6.3.1.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for imputation set 1 was the preliminary analysis weight, PANALWT. As with the 1999–2011 surveys, the 2012 survey implemented automatic routing of the respondent through the questionnaire based on the respondent's answers, thereby skipping over (i.e., not asking the respondents) specific questions. Within each drug module, one (e.g., for marijuana) or multiple (e.g., for hallucinogens) questions were asked in order to establish whether the respondent had ever used the drug in question during his or her lifetime. For more information on gate questions and multiple gate questions see section 6.2.1.

For an individual to be considered a lifetime use item respondent, he or she must have complete data for all of the drug module gate questions: cigarettes; cigars; chewing tobacco; snuff; pipes; alcohol; marijuana; cocaine; crack; heroin; inhalants; LSD; PCP; Ecstasy; hallucinogens other than LSD, PCP, and Ecstasy; OxyContin®; pain relievers other than OxyContin®; tranquilizers; methamphetamine; stimulants other than methamphetamine; and sedatives. See Table D.3 in Appendix D for details of the covariates used in the RP models for these variables.

6.3.1.2 First Prediction Step (Lifetime Smokeless Tobacco Use)

Many respondents who indicated lifetime use of smokeless tobacco seemed to be confused regarding the difference between chewing tobacco (chew) and snuff, as was demonstrated by their responses to questions regarding specific brands. For example, many respondents who indicated use of chewing tobacco entered a snuff brand, such as Copenhagen™, when asked about the specific brand of chew they used. As a result, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted, rather than individual models for chew and snuff. The probability of lifetime smokeless tobacco use was modeled for item respondents within each age group, using the nonresponse-adjusted weights. SUDAAN's RLOGIST procedure was used to perform dichotomous logistic regression[83] to determine the parameter estimates and probability of use for both respondents and nonrespondents.

6.3.1.3 First Provisional Hot-Deck Step (Lifetime Smokeless Tobacco Use)

In order to use lifetime usage of a given drug as a covariate for a drug later in the sequence, it was necessary to create temporary imputed values in cases where the original lifetime usage indicator was missing. Lifetime indicators for both chew and snuff were used as covariates for later models, so it was necessary to create these provisional values. In the first provisional hot-deck step, matching was done on a single predicted mean from the PRD step, but missing values for both chew and snuff were replaced with the values from a donor within this neighborhood.

[83] SAS®-callable SUDAAN® was used to fit all dichotomous and polytomous logistic regression models. Details about the logistic regression model and additional references can be found in RTI International (2008). SAS software is a registered trademark of SAS Institute Inc. SUDAAN is a registered trademark of Research Triangle Institute.

If possible, donors and recipients were required to be from States with the same level of smokeless tobacco usage (State rank[84]), where the level of usage was defined in terms of the weighted proportion of a given State's residents who were lifetime users of the drug.[85] An additional likeness constraint required the donor to match the recipient on any nonmissing lifetime use indicators for child drugs. For example, if the lifetime use indicator for overall smokeless tobacco was missing, but the recipient was known to be a lifetime nonuser of snuff, then the donor must also have been a lifetime nonuser of snuff. If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) both the State-rank and child lifetime drug indicator constraints were removed, and the delta constraint was reapplied.

No logical constraints were placed on the neighborhoods for any of the lifetime usage indicators. Even in the case of smokeless tobacco where more than one substance was associated with a single predicted mean, leading to a multivariate assignment of provisional imputed values, no logical constraints were necessary.

6.3.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Drugs

PRD models and provisional hot-deck steps were completed in a manner similar to that described above for cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine and crack, and heroin, with the following deviations:

- For cigars, pipes, alcohol, inhalants, marijuana, tranquilizers, and sedatives, only one substance was associated with the predicted mean from the modeling stage. In these cases, the donor directly supplied the overall drug use value rather than providing values for child drugs that were then combined into a final usage measure as was the case for smokeless tobacco.
- Because cocaine and crack were in two separate back-to-back modules in the 2012 NSDUH questionnaire, separate models were fitted for the two substances. However, crack is a type of cocaine, so donors for the two substances were obtained using a single neighborhood with multivariate matching.[86] This was true regardless of whether the item nonrespondent was missing only crack, only cocaine, or both crack and cocaine. Once the neighborhood was defined, missing values for crack and/or cocaine were replaced with the values from one donor within this neighborhood.
- For hallucinogens, pain relievers, and stimulants, predicted probabilities were calculated for the parent drugs, and these probabilities were used to determine neighborhoods for each group of drugs. Lifetime usage indicators were assigned for LSD, PCP, Ecstasy, and "other" hallucinogens; OxyContin® and "other" pain relievers; and methamphetamine and "other" stimulants. Lifetime usage indicators for

[84] See Section 3.5 for a general discussion of State-rank variables.

[85] Those with a missing lifetime use indicator for the drug were treated as lifetime nonusers.

[86] This provisional hot-deck program actually uses a multivariate delta constraint, but the distance from donor to recipient is based only on the predicted probability of lifetime use of cocaine. This was done to avoid the complexity of the Mahalanobis distance calculation. Strictly, this is neither univariate nor multivariate matching.

the parent drugs were created later by combining the constituent parts, including the "other" group of substances.

- Heroin did not undergo a provisional imputation step, because it was the last variable in the imputation set.

6.3.1.5 Final Hot-Deck Step

Tables E.20 through E.23 in Appendix E provide details on the final hot-deck step for the lifetime use indicators. Although the predictive mean vector could be large if several indicators were missing, the hot-deck step included fairly simple constraints. Only one logical constraint was used for lifetime use of pain relievers: those item nonrespondents who were known to have used pain relievers, but both their OxyContin® and "other" pain reliever indicators were missing, were required to have a donor who was a lifetime user of pain relievers. This pattern of nonresponse occurs either when respondents respond affirmatively to PR04 (lifetime use of one of the drugs appearing on the card, which includes OxyContin®) but fail to select any drugs from the card in PR04A or when respondents refuse to respond to each individual gate question (that specifically ask about certain pain relievers) but then respond affirmatively to the probe question PRREF.

No final imputation-revised variables indicating lifetime usage alone were created, because this information was recorded in the final imputation-revised recency-of-use variables. Imputation indicators also were not created, though temporary variables indicating that lifetime usage was imputed were maintained to inform the creation of the recency-of-use imputation indicators.

6.3.2 Imputation-Revised Cigarette Recency and Frequency of Use (Imputation Set 2)

As indicated in Table 6.5, the second set of drug use variables to undergo imputation were the cigarette recency and 30-day frequency variables. The multiple RP/multiple PRD type of PMN was used to process these variables.

6.3.2.1 Sequence of Imputation

Because recency-of-use and frequency-of-use variables for a given drug were in the same imputation set, the calculation of predicted means for the frequency-of-use variables required the item nonrespondents to be identified as provisional past month and/or past year users. For this reason, cigarette recency was modeled prior to 30-day frequency, and provisional imputations were performed to allow for the identification of provisional past month users of cigarettes.

6.3.2.2 First Response Propensity Step (Cigarette Recency of Use)

The input to the weight adjustment model in the response propensity step for imputation set 2 was the preliminary analysis weight, PANALWT. To impute for missing recency-of-use values for cigarettes, it was first necessary to define the domain within each of the three age groups. Using the imputation-revised lifetime indication of use, the file was reduced to lifetime cigarette users. Among these lifetime users, item respondents and nonrespondents were identified across recency-of-use and 30-day frequency-of-use measures. If a valid response was

provided for each drug use measure, the person was deemed an item respondent for cigarettes. Otherwise, he or she was an item nonrespondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.2.3 First Prediction Step (Cigarette Recency of Use)

Using the adjusted weights, the probability of selecting each cigarette recency-of-use category was modeled within each age group using polytomous logistic regression. SUDAAN's MULTILOG procedure was used to estimate the parameters from the appropriate logistic model from which predicted probabilities for each of the cigarette recency categories were calculated for both item respondents and item nonrespondents. The four recency categories were the following:

1. past month;
2. past year, not past month;
3. past 3 years, not past year; and
4. lifetime, not past 3 years.

6.3.2.4 First Provisional Hot-Deck Step (Cigarette Recency of Use)

In order to define the domain for the cigarette 30-day frequency-of-use variable, it was necessary to create temporary imputed values in cases where the original cigarette recency value was missing. In order to save time and resources and because the imputation was only provisional, a univariate matching procedure was implemented. The only predicted mean used was the predicted probability of past month use, because past month use was the most critical measure of recency of cigarette use.

If possible, donors and recipients were required to be from States with the same level of usage of a given drug (State rank; see Section 3.5), where the level of usage was defined in terms of the weighted proportion of a given State's residents who had used cigarettes in the past month.[87] If insufficient donors were available within these constraints, they were loosened in the following order: (1) the delta constraint was removed, and (2) donors and recipients were no longer required to be from States with similar usage levels.

The only logical constraints placed on the neighborhoods involved cases where a general recency category was available for a respondent and imputation was required to determine the specific recency categories. The general recency categories that appeared are shown in Table 6.6. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood.

[87] Those individuals whose past month use status was unknown were treated as if they were not past month users.

Table 6.6 General Incomplete Recency Categories for Cigarettes

General Incomplete Recency Category	Allowable Specific Recency Categories
Lifetime	1. Past month 2. Past year but not past month 3. Past 3 years but not past year 4. Lifetime but not past 3 years
Past Year	1. Past month 2. Past year but not past month
Lifetime, Not Past Year	1. Past 3 years but not past year 2. Lifetime but not past 3 years
Lifetime, Not Past Month	1. Past year but not past month 2. Past 3 years but not past year 3. Lifetime but not past 3 years
Lifetime, Not Past Month but within Past 3 Years	1. Past year but not past month 2. Past 3 years but not past year
Past 3 Years	1. Past month 2. Past year but not past month 3. Past 3 years but not past year

6.3.2.5 Second Response Propensity Step (Cigarette 30-Day Frequency)

The input to the weight adjustment model in the response propensity step for imputation set 2 was the preliminary analysis weight, PANALWT. The modeling of cigarette 30-day frequency followed that of recency. The file was first reduced to the domain, which was past month cigarette users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the cigarette recency imputations. To be an item respondent, the individual had to have provided valid responses to both the cigarette recency and 30-day frequency measures. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of cigarettes. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.2.6 Second Prediction Step (Cigarette 30-Day Frequency)

As stated in the previous section, only past month users of cigarettes were used to build the 30-day frequency-of-use model. The response variable of interest in the 30-day frequency-of-use models, prior to a normalizing transformation, was the proportion of the days in a month (30 days) on which a respondent used cigarettes. The range of values for the proportion was from (greater than) 0 to 1. Hence, to model 30-day frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i + 0.5\right)/\left(N - Y_i + 0.5\right)\right],$$

where Y_i was the observed 30-day frequency for respondent i and N was 30, the total number of days in the month that the respondent could have used the substance. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/(1-P_i)\right],$$

where P_i was defined as the proportion of days in the past month on which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[88] Using the adjusted weights, a linear univariate regression model was then fitted using SUDAAN software for the log-transformed variable Y_i within each age group.

For cigarettes, the empirical distribution for 30-day frequency of use was in fact a mixture distribution, with a positively skewed distribution from 1 to 29 and a spike at 30. This substance was modeled using two separate models. One was a logistic model for daily use versus nondaily use among past month users. For the nondaily past month users (i.e., those who had used between 1 and 29 days), the model described above was used. In this case, the response variable in a linear regression model was a logit of the proportion of the period (30 days) during which a respondent used the substance. Both the predicted probability of daily use and the logit of the proportion of the month used (assuming nondaily use) were used as predicted means in the subsequent hot-deck step. The logit was back-transformed into a proportion before use in the hot-deck step.

6.3.2.7 Final Hot-Deck Step (Cigarette Recency and 30-Day Frequency)

The full predictive mean vector for cigarettes contained probabilities associated with several of the recency-of-use categories, a probability of daily use, and a predicted probability of use on a given day in the past month. Each element in the full vector of predicted means was adjusted so that all elements were conditioned on the same usage status whenever possible. The elements in the predictive mean vector that could have potentially resulted are shown in Table 6.7, with the assumption that only the lifetime usage is known. If other information about the recency of use is known (e.g., past year user), the predictive mean vector is adjusted accordingly. The portion of the full predictive mean vector used to determine the neighborhood for a particular item nonrespondent was dependent on the pattern of missingness for that item nonrespondent. If partial information was available regarding recency of use, then that information was used to adjust the recency-of-use probabilities. The portions of the full predictive mean vector used for each missingness pattern, with accompanying adjustments, are provided in Table E.28 in Appendix E. The Mahalanobis distance was then calculated using only the portion of the predictive mean vector that was associated with the given missingness pattern, with elements appropriately adjusted. The likeness and logical constraints applied to each missingness pattern are also available in Table E.28.

[88] If the respondent was a daily user of the substance, then $\log[(Y+0.5)/(N-Y+0.5)] \approx \log[(N+0.5)/0.5]$ with N = 30 so that it was defined for all respondents. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

Table 6.7 Elements of Full Predictive Mean Vector: Cigarettes

Drug Use Measure and Category of Interest	Predicted Mean
Recency of Use, Past Month Use[1]	*P*(past month user | lifetime user)
Recency of Use, Past Year but Not Past Month Use[1]	*P*(past year but not past month user | lifetime user)
Recency of Use, Past 3 Years but Not Past Year Use[1]	*P*(past 3 years but not past year user | lifetime user)
30-Day Frequency of Use for Nondaily Users over Past 30 Days	*P*(use on a given day in the month | past month user, not a daily user) × *P*(not a daily user | lifetime user) × *P*(past month user | lifetime user)[2]
Daily User over Past 30 Days	*P*(daily user | past month user) × *P*(past month user | lifetime user)[2]

[1] The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.

[2] Interpreting the proportion of the month used as a probability of use on a given day in the month assumed that the probability of use on each day in the month was equal, which was not true.

6.3.2.8 Final Variables (Cigarette Recency and 30-Day Frequency)

The final imputation-revised recency-of-use and 30-day frequency variables were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the drug (CIG) and the final two letters identified the measure (RC = recency of use, FM = 30-day frequency). Each IR variable was accompanied by an imputation indicator with a prefix II instead of IR. The levels for the imputation indicators were the standard levels used for all imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user).

6.3.2.9 Recodes for Additional Analyses

From the final imputation-revised recency-of-use variable, three dichotomous indicator variables were created to indicate cigarette use in the lifetime (CIGFLAG), past year (CIGYEAR), or past month (CIGMON).

6.3.3 Imputation-Revised Cigarette Age at First Use (Imputation Set 3)

As indicated in Table 6.5, the third imputation set consisted of the cigarette age-at-first-use variable. Unlike the recency and 12-month frequency-of-use variables, age at first drug use was not statistically imputed in the surveys prior to 1999. Instead, missing values were excluded from subsequent analyses. However, as with the 30-day frequency, missing age-at-first-use values have been replaced using imputation since the 1999 survey. Also, recent drug initiates (i.e., those whose current age was equal to or 1 year greater than the reported age at first use) were asked the year and month of their first use. To have this information for all users, both missing year and missing month of first use for less recent initiates (and recent initiates who did not report year and month of first use) were replaced by assigning values consistent with the respondent's current age, interview date, imputation-revised age at first use, and imputation-revised recency and frequency variables. To have complete date-of-first-use information, day of

first use was randomly assigned for all users. The combined data gave the respondent's age at first use along with the date of first use.

6.3.3.1 Response Propensity Step (Cigarette Age at First Use)

The input to the weight adjustment model in the response propensity step for imputation set 3 was the preliminary analysis weight, PANALWT. To impute for missing age at first use for cigarettes, it was necessary to define the eligible population. Using the imputed recency of use, the files were reduced to lifetime users of cigarettes. If a valid response was provided for the age-at-first-use measure,[89] the person was deemed an item respondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.3.2 Prediction Step (Cigarette Age at First Use)

The response variable in the model for age at first use, before a normalizing transformation, was the age at first use as a proportion of the current age. The numerator in this proportion was an integer representing age at first use. However, because this integer was in fact a truncated version of the real age at first use, the value was made continuous by adding a random component between 0 and 1. Hence, expressing the proportion as $P_i = Y_i/N_i$, the numerator was given as

$$Y_i = \textit{Age at First Use}_i + \textit{Uniform(0,1) random number}.$$ [90]

The denominator in the proportion was the total age. The true age was known, based on the interview date and birth date. Expressing it in years rather than days required dividing by the number of days in the year:

$$N_i = (\text{Interview Date} - \text{Birth Date} + 1)/365.25 .$$

After a weight adjustment, the empirical logit transformation was used as the response variable in a weighted linear univariate regression:

$$\log\left[\left(Y_i + 0.5\right)/\left(N_i - Y_i + 0.5\right)\right].$$

This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

which was not used, because it might be unstable for respondents who started using at their current age.

---

[89] Respondents who reported an age at first use of 1 or 2 were treated as item nonrespondents in the response propensity and prediction steps, because of the implausibility of such a young age at first use. In the hot-deck step, their response was left unchanged, but they were not allowed to be donors.

[90] In the event that the age at first use was equal to the age, Y_i was constrained so that it was equally likely to be anywhere on the interval [*Age at First Use*$_i$, N_i]. Thus, Y_i was prevented from being greater than N_i.

One unusual covariate used in the PRD model for cigarette age at first use was a modified 30-day frequency variable for cigarettes. It was defined as follows:

$\mathbf{new30_i}$	=	0	if respondent *i* did **not** use cigarettes in the past month
	=	30-day frequency	if respondent *i* used cigarettes in the past month

Naturally, the full model for age at first use did not include the lifetime indicator for the drug in question, because the model was built on cigarette users. A summary of the starting and final models can be found in Tables D.4, D.5, and D.6 in Appendix D.

From the final model, a predicted value (based on the *Y* variable) was calculated for each cigarette user, which was then back-transformed to produce a predicted cigarette age at first use.

6.3.3.3 Hot-Deck Step

The imputation-revised cigarette age-at-first-use assignment was conducted using a single predicted mean: the predicted age at first use. Tables E.58 through E.60 provide a complete list of likeness and logical constraints applied to the cigarette age at first use imputations. The likeness constraints for age at first use were more stringent than those for the other drug use measures. Therefore, it was often necessary to loosen the constraints. Once these likeness constraints were removed, some complex logical constraints remained, based on the interview date, the birth date, and imputation-revised recency and frequency values.

6.3.3.4 Date-of-First-Use Assignments

After the age-at-first-use imputations, all lifetime users of cigarettes had nonmissing age-at-first-use values. Using this age at first use (AFU), users were assigned values for year, month, and day of first use. Recent initiates, or those respondents whose AFU was within 1 year of his or her age, were asked for their year of first use (YFU) and month of first use (MFU). The day of first use (DFU) was not collected in the questionnaire and was missing for all respondents. The YFU, MFU, and DFU data contained four patterns of missingness:

Pattern 1: Recent initiates: missing day of first use only;

Pattern 2: Recent initiates: missing month/day of first use;

Pattern 3: Recent initiates: missing year/month/day of first use; and

Pattern 4: Less recent initiates: missing year/month/day of first use.

For each missingness pattern, upper and lower bounds on the date of first use (i.e., the earliest possible date of first use and the latest possible date of first use) were determined. Once the earliest and latest possible dates of first use were determined, a day was randomly selected from this interval. The imputation-revised month/day/year values were then extracted from this date of first use.

6.3.3.4.1 Missingness Pattern 1

In this missingness pattern, a recent initiate provided all the information asked by the questionnaire (i.e., both the MFU and YFU). However, to obtain a complete date of first use, a

DFU also was needed. Thus, a DFU was randomly assigned, given the respondent's month and year of first use, in a way that was consistent with both the 30-day frequency/recency and age at first use. Below is a brief description of the process used to obtain a date of first use in such cases. The imputed YFU, MFU, and DFU were extracted from the date, as defined below:

> *Final date of first use = Earliest possible date* + [(*Days between earliest and latest date*) × (a random number generated from a Uniform (0,1) distribution)],

where

> *Days between earliest and latest date = Latest possible date – Earliest possible date + 1;*
>
> *Earliest possible date = maximum [(AFUth birthday), (first day of the month indicated by MFU/YFU)]; and*
>
> *Latest possible date =*

- minimum [(Interview date – 30-day frequency + 1), (1 day before the (AFU + 1)th birthday)], *if recency* = 1;
- minimum [(Interview date – 30), (1 day before the (AFU + 1)th birthday)], *if recency* = 2; and
- minimum [(Interview date – 1 year), (1 day before the (AFU + 1)th birthday)], *if recency* = 3.

Note that it is impossible for recent initiates to have recency = 4 (lifetime but not past 3 years). Recent initiates had to have begun using the drug no earlier than their (AFU)th birthday. Because AFU = current age, or AFU = current age – 1, their (AFU)th birthday was within the past 2 years. Respondents who had begun using the drug within the past 2 years must logically have last used the drug within the past 2 years, and therefore could not have had recency = 4.

In rare cases, the *earliest possible date* was set to 29 days before the interview. This occurred for respondents meeting all of the following conditions:

1. The *latest possible date* was within 29 days of the interview.
2. The *earliest possible date* determined by the above rule was within a year of the interview.
3. The recency = 1.
4. The 12-month frequency = 30-day frequency (if applicable), or the 12-month frequency = 1.

Logically, all the lifetime usage of the drug for these respondents occurred in the past 30 days (including the interview date). The first condition ensures that the application of this rule does not cause an inconsistency. The second condition implies that the drug was not used by these respondents more than 1 year ago. The third and fourth conditions imply that the drug was

not used by these respondents in the interval (1 year before the interview, 1 month before the interview). Therefore, these respondents did not use the drug more than 1 month ago. All their lifetime use must have occurred in the past month.

6.3.3.4.2 Missingness Pattern 2

The second missingness pattern occurred when a recent initiate provided his or her YFU but did not provide an MFU. In such cases, a month and day were randomly assigned that were consistent with both the respondent's frequency/recency and with the age-at-first-use range. The imputed MFU and DFU were derived in the same manner as the date of first use in Missingness Pattern 1, except with the following changes:

- For the *earliest possible date*, replace "first day of the month indicated by MFU/YFU" with "January 1st of the YFU."
- For the *latest possible date*, replace "last day of the month indicated by MFU/YFU" with "December 31st of the YFU."

6.3.3.4.3 Missingness Pattern 3

Similar to Missingness Pattern 2, the third missingness pattern occurred when recent initiates provided neither an MFU nor a YFU value. In these cases, the year/month/day of first use were randomly assigned from a uniform distribution in a way that was consistent with both the cigarette 30-day frequency/recency and the age at first use. Again, the imputed YFU, MFU, and DFU were derived in the same manner as described in Missingness Pattern 1.

6.3.3.4.4 Missingness Pattern 4

The fourth missingness pattern occurred when the respondent reported, or was imputed to, an age at first use at least 2 years less than his or her age. This case is analogous to data prior to the 1999 survey, where month and year of first use were not asked in the questionnaire. In this missingness pattern, the 30-day frequency was immaterial to the final date of first use because the respondent could not have begun using in the past month:

Earliest possible date = AFUth birthday; and

Latest possible date =

- 1 day before the (AFU + 1)th birthday, *if recency* < 4; or
- minimum [(Interview date – 3 years), (1 day before the (AFU + 1)th birthday)], *if recency* = 4.

6.3.3.5 Final Age and Date-of-First-Use Variables

As with all other imputation-revised variables, the final imputation-revised date-of-first-use variables were identified with the prefix IR, followed by a six-letter identifier, where a three-

letter code identified the drug[91] and the final three letters identified the measure (AGE = age at first use, YFU = year of first use, MFU = month of first use, DFU = day of first use). Each IR variable was accompanied by an imputation indicator with the requisite II prefix. The levels for the imputation indicators were the standard levels used for all imputation-revised variables: 1 = questionnaire data, 2 = logically assigned, 3 = statistically imputed, and 9 = legitimate skip (not a lifetime user). Because survey respondents are not asked for the specific day on which they first used the drug of interest, all respondents in the domain receive IIxxxDFU = 3. Also, as indicated above, only recent initiates are asked for the year and month of first drug use. Subsequently, these questions have high rates of nonresponse because of the skip logic embedded in the questionnaire, as all other individuals in the domain require imputation for their year and month of first use.

6.3.4 Imputation-Revised Age at First Daily Cigarette Use (Imputation Sets 4 and 5)

In addition to age at first use, the cigarettes module also included a question asking for the respondent's age at first daily cigarette use, where a daily user was defined as someone who reported having at some time smoked cigarettes every day for a period of at least 30 days. Imputation procedures for age at first daily cigarette use were similar to age at first use, with some key exceptions as discussed below.

One such exception involved the domain of the age-at-first-use variable. Whereas the age-at-first-use question was asked of all cigarette users, the age-at-first-daily-use question was asked of only daily users. The "daily use" indication came from two sources. If a respondent answered either the 30-day frequency or estimated 30-day frequency with a "30," or if the respondent had a "yes" value for the edited variable associated with the "ever daily used" question (CIGDLYMO), then he or she was considered a daily user. For more information about CIGDLYMO, see Section 6.2.7. The "ever daily used" question (CIGDLYMO) can be thought of as a lifetime "child" drug to the "parent" lifetime cigarette use question (CIGEVER). However, anyone who answers the 30-day frequency or estimated 30-day frequency with a "30" is automatically skipped out of this question. Therefore, it is necessary to obtain the imputation-revised cigarette 30-day frequency (IRCIGFM) prior to imputing the lifetime-daily-cigarette-use variable (IRCDULF) so that it is not included with the other lifetime drug indicators as part of imputation set 1. Instead, as indicated in Table 6.5, the age at first daily cigarette use actually contains two separate imputation sets. Imputation set 4 includes the lifetime indicator of daily cigarette use and imputation set 5 includes the age-at-first-daily-cigarette-use variables. At this stage in the process, there should be no missing responses to the 30-day frequency question, which were imputed as part of imputation set 2 as discussed above. Daily users, based on 30-day frequency, should be either known (based on a response in the survey) or imputed. However, responses for the ever-daily-used question (CIGDLYMO) could still be missing, and, therefore, it was first necessary to impute these values to define the domain for the age-at-first-daily-use variable.

[91] Exceptions to this rule occurred with marijuana and cigarette daily use. For historical reasons, marijuana contained a two-letter code (MJ). Marijuana variables therefore ended with a five-letter identifier rather than a six-letter identifier. The code for cigarette daily use was CDU, which differed from the general cigarette code of CIG. Details about cigarette daily use are provided in Section 6.3.4.

6.3.4.1 Response Propensity Step (Ever-Daily-Used Cigarettes)

The input to the weight adjustment model in the response propensity step for imputation sets 4 and 5 was the preliminary analysis weight, PANALWT. To impute for missing values in the ever-daily-used variable, it was necessary to define the domain: lifetime users of cigarettes who had an imputation-revised 30-day frequency[92] of fewer than 30 days (includes legitimate skip codes for lifetime but not past month users). If a valid response was provided in the ever-daily-used variable, the person was deemed an item respondent. See Tables D.4, D.5, and D.6 in Appendix D for details of the covariates used in the RP models for this variable.

6.3.4.2 Prediction Step (Ever-Daily-Used Cigarettes)

After the weights were adjusted, the ever-daily-used variable was modeled using weighted logistic regression in SUDAAN. The predicted mean from this model was the predicted probability of ever smoking cigarettes daily.

6.3.4.3 Hot-Deck Step (Ever-Daily-Used Cigarettes)

The predicted mean from the prior step was used in a straightforward hot-deck step, which is summarized in Tables E.24 and E.25.

6.3.4.4 Hot-Deck Step (Age at First Daily Cigarette Use)

Instead of separately modeling age at first daily cigarette use, the predicted means from the age-at-first-cigarette-use models were used to determine neighborhoods. The imputation-revised age-at-first-daily-use assignment was conducted using univariate matching and univariate assignment.

All the logical constraints applied to age at first cigarette use were also applied to age at first daily cigarette use. Besides those logical constraints, an additional logical constraint was applied specifically to age at first daily cigarette use. If the cigarette age at first use was not missing for a recipient with a missing age at first daily use, the donors were prevented from having an age at first daily use earlier than the preexisting cigarette age at first use. This constraint was applied as daily cigarette users constitute a subset of all cigarette users. Therefore, daily use of cigarettes can be thought of as a child drug for cigarettes, with a lifetime indicator and an age-at-first-use measure but no recency or frequency. This association required that these variables remain internally consistent for each respondent.

The likeness constraints were nearly identical to those used for cigarettes age at first use. There was only one difference: an additional step was employed if no donor was found after loosening all of the likeness constraints. In particular, if the age at first use and age at first daily

[92] The imputation-revised 30-day frequency included responses from the 30-day frequency question (CG07), as well as the estimated 30-day frequency question (CG07DKRE).

use were both initially missing, the imputed age at first use was set back to missing and reimputed simultaneously with the age at first daily use so that they were mutually consistent.[93]

Subject to these constraints, the age-at-first-daily-use value of the randomly selected donor was then assigned to the recipient.

6.3.4.5 Assignment of Date of First Daily Cigarette Use

After the imputation-revised age at first daily cigarette use was created, all daily cigarette users had a valid age at first daily cigarette use. From this age, a year/month/day of first daily use was assigned. The date assignment procedure was identical to the procedure described in Section 6.3.3.4 with the following exception. In the setting of the *earliest possible date* for daily cigarette use, the overall cigarette date of first use was used as an additional bound. This was done for cigarettes and other substances to ensure that the child drug's date of first use was never earlier than the parent drug's date of first use.

6.3.4.6 Final Variables

The final imputation-revised date-of-first-daily-cigarette-use variables were named in the same manner as described in Section 6.3.3.5. However, the three-letter identifier for cigarette daily use was CD2. As with the general cigarette use variables, each IR variable was accompanied by an imputation indicator with a prefix II instead of IR.

6.3.5 "Other" Drugs Recency and Frequency

Imputation for the parent and child recency and frequency variables for imputation sets 6, 8, 10, 11, 13, 15, 17, 19, 21, 23, 25, 27, and 29 in Table 6.5 (smokeless tobacco, cigars, pipes, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine, and heroin, respectively) was done in a manner similar to that described above for cigarettes. The following deviations from the process described for cigarettes applied to these "other" drugs.

The order of imputation for smokeless tobacco and cigars was identical to that for cigarettes, with recency of use being modeled first, followed by 30-day frequency. However, not all imputation sets included the same variables. Alcohol, inhalants, marijuana, hallucinogens, cocaine and crack, and heroin also included a measure for 12-month frequency of use. For these drugs, imputation proceeded in the following order: recency of use, 12-month frequency of use, and 30-day frequency of use. For a given drug, this ordering allowed recency of use to be included in the model for 12-month frequency of use and allowed 12-month frequency of use to be included in the model for 30-day frequency. Further, this ordering allowed the provisional recency of use to define the domains for the frequency measures. Alcohol also had a measure for binge drinking frequency, which was modeled after the 30-day frequency-of-use variable so that the provisionally imputed value could be used as a covariate in the binge drinking frequency model. For pain relievers, tranquilizers, stimulants, and sedatives, the respondents were not asked to report their 30-day frequency of use. For these imputation sets, recency of use was

[93] Though it has occurred in prior years, the situation where no donors were available, even after loosening all constraints, did not occur in the 2012 NSDUH.

completed first, followed by the 12-month frequency-of-use variable. For pipes, the respondents were only asked about their most recent use, and no information was collected regarding frequency of use in the past year or month. Therefore, only the recency-of-use variable required modeling and imputation.

6.3.5.1 Recency of Use

6.3.5.1.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the parent and child recency-of-use variables was the preliminary analysis weight, PANALWT. Similar to cigarettes, the eligible population for the recency-of-use models included all lifetime users of the drug of interest as identified by the imputation-revised lifetime drug use variables. However, the identification of respondents and nonrespondents for the purpose of imputation differed from drug to drug depending on the information collected in the questionnaire. In general, an individual had to have provided a valid response to all variables included in the imputation set to be classified as a respondent. If the imputation set included both "parent" and "child" drugs (Table 6.1), then this requirement extended across all observed measures of drug use. For example, to be classified as a respondent for the hallucinogens imputation set, a valid response must be provided for the overall hallucinogens recency of use, 12-month frequency of use, and 30-day frequency of use, as well as the recency of use for LSD, PCP, and Ecstasy questions. See Tables D.7 through D.45 in Appendix D for details of the covariates used in the RP models for these variables.

6.3.5.1.2 Prediction Step

Only cigarettes, cigars, chewing tobacco, and snuff included a recency category for past 3 years but not past year. For all other drugs except pipes, the outcome variable was a three-level categorical variable with the following levels:

1. past month
2. past year, not past month
3. lifetime, not past year

For pipes, the outcome variable had only two levels:

1. past month
2. lifetime, not past month

Because cigarettes were the first recency/frequency imputation set, it was not possible to include the recency information for other drugs as covariates in the PRD model. However, for drugs other than cigarettes, recency-of-use covariates for cigarettes, alcohol, and marijuana replaced the lifetime indicators where applicable. For example, the PRD model for alcohol included recency indicators for cigarettes but only included the lifetime usage indicator for marijuana because this drug comes later in the sequence.

6.3.5.1.3 Provisional Hot-Deck Step

For certain cases, a general recency category was assigned during the editing process, and the specific recency was then determined during imputation. However, the categories available for both the general recency and the specific recency varied from drug to drug depending on the number of levels included in the recency-of-use measure. The allowable general and specific recency categories for cigarettes are shown in Table 6.6. The same categories apply to cigars, chewing tobacco, and snuff. For all other drugs except pipes, the only general incomplete recency categories that were applicable were lifetime and past year (the first two rows). For pipes, only the lifetime category was applicable. Logical constraints ensured that only donors with allowable specific recency categories were included in the neighborhood of potential donors.

Occasionally, more than one substance was associated with a single predicted mean, leading to a multivariate assignment of imputed values. However, for the provisional imputed values, a multivariate assignment was necessary only if the substances associated with a single predicted mean were of equal standing. This occurred with smokeless tobacco, which consists of chewing tobacco and snuff. No provisional imputed values were determined for substances that were a subset of the substance associated with the predicted mean (parent/child drugs). Examples of such situations included cocaine (parent) and crack (child); pain relievers (parent) and OxyContin® (child); stimulants (parent) and methamphetamine (child); and hallucinogens (parent) and LSD, PCP, and Ecstasy (child).

As with lifetime use, one model for smokeless tobacco (a combination of the chew and snuff responses) was fitted rather than individual models for chew and for snuff. The nearest neighbor hot-deck neighborhood was then based on the predicted probability of past month use of smokeless tobacco. Missing recency-of-use values for chew and/or snuff were replaced with the (provisional) values from a donor within this neighborhood. The provisional recency of use for smokeless tobacco was obtained by combining the recency-of-use information from chew and snuff.

6.3.5.1.4 Hierarchical Modeling

For certain drugs, the proportion of users who were past year users was quite small when compared with the total number of lifetime users. The lopsided distributions[94] for these drugs caused convergence problems when fitting polytomous logistic models. This problem occurred with the following set of drugs that were either rare overall or were rare within one or more age groups: inhalants, hallucinogens, sedatives, stimulants, tranquilizers, and heroin. To alleviate this problem, the single polytomous logistic model was replaced with two dichotomous logistic models[95] that were fit hierarchically.

---

[94] A "lopsided distribution" in the context of recency of use is where, among the categories of past month use, past year but not past month use, and lifetime not past year use, only a small minority of respondents gave a response of "past month use."

[95] The set of covariates used for these dichotomous logistic models were the same as those for logistic modeling given earlier in this section.

As with the polytomous logistic model, the first dichotomous logistic model was fit among lifetime users, but the past month and past year but not past month categories in the response variable were collapsed into a single level. In a similar manner to other recency-of-use models, respondents' weights were adjusted so that they represented all lifetime users. The predicted probability of past year use given lifetime use was obtained from this model.

The second model was limited to past year users, where the response variable had two levels: past month and past year but not past month users. For the second model, respondents' weights were adjusted so that they represented all past year users. In order to do this, it was necessary to completely define the domain of past year users. Missing values were provisionally imputed to past year or not past year use by randomly allocating the response using the predicted means from the first model.

From the two dichotomous logistic models, both the probability of past month use and the probability of past year but not past month use were obtained and used in the provisional hot-deck program for recency. Once the predicted means were determined from the two models, a single vector of predicted means conditional on lifetime usage, as with the polytomous logistic models, was determined as follows:

$$P(\text{past month use} \mid \text{lifetime use}) = P(\text{past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}), \text{ and}$$

$$P(\text{past year, not past month use} \mid \text{lifetime use}) = P(\text{past year, not past month use} \mid \text{past year use}) \times P(\text{past year use} \mid \text{lifetime use}).$$

6.3.5.2 12-Month Frequency of Use

The modeling of 12-month frequency sequentially followed that of recency of use for alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, sedatives, cocaine and crack, and heroin.

6.3.5.2.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the 12-month drug frequency-of-use variables was the preliminary analysis weight, PANALWT. The eligible population for the imputation of 12-month frequency of use was past year users of the drug in question (as defined by the provisional recency of use). The item response indicator and the response propensity adjustment were defined among the past year users of each drug. Item respondents were defined using the same criterion as was used in the recency-of-use imputations. Namely, the respondent had to have a valid response to all of the applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past year users of the drug. See Appendix D for details of the covariates used in the RP models for these variables.

6.3.5.2.2 Prediction Step

As indicated in the previous section, only past year users of the drug of interest were used to build the 12-month frequency-of-use model. The response variable of interest in the 12-month

frequency-of-use models for most respondents, prior to a normalizing transformation, was the proportion of the days in a full year (365.25) on which a respondent used a particular drug. For example, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/365.25. Some respondents, however, started using the drug within the past year. If they responded to the month-at-first-use question, the difference between the month of first use and the date of the interview indicated the total time period during which they could have been using drugs.[96] If the date of the interview was July 10, for example, and the month of first use was March of the same year, the maximum period during which the respondent could have used is the number of days between March 1 and July 10 (inclusive), or 101. Thus, if a respondent entered a 12-month frequency of 100, the (untransformed) response variable of interest would be 100/101 instead of 100/365.25. The range of values for the proportion was from (greater than) 0 to 1. Hence, in order to model 12-month frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i + 0.5\right)/\left(N_i - Y_i + 0.5\right)\right],$$

where Y_i is the observed 12-month frequency for respondent i and N_i is the total number of days in the year that respondent i could have used the substance. This transformation is nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

where P_i is defined as the proportion of days in the past year in which respondent i used the drug. The standard logit transformation was not used because it was not defined for daily users.[97] Using the adjusted weights, a linear univariate regression model using SUDAAN software was then fitted for the log-transformed variable Y_i within each age group.

- **Covariates:** Because the 12-month frequency models were limited to past year users, only two recency categories could have resulted: past month use and past year but not past month use.[98] Hence, recency of use for the drug being modeled was represented as a covariate in the 12-month frequency-of-use model by a single indicator variable representing these two categories. Imputation-revised recency of use for other drugs was used if available. If the missing values for a given drug's recency of use had not yet been imputed, a single covariate was used that indicated lifetime usage of that drug. To control for State variations in drug use, the State-rank groups defined for the recency-of-use imputations were included as covariates in the 12-month frequency-

[96] If a respondent initiated use in the past year (according to his or her age-at-first-use response), but did not answer the month-at-first-use question, then the maximum period the respondent could have been using drugs was assumed to be 365.25 because no other information was available.

[97] If the respondent was a daily user of the substance, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$ with N = 365.25 so that it was defined for all respondents. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

[98] For item nonrespondents, where parameter estimates were used to determine predicted means, past year use was defined based on a provisional imputation.

of-use models.[99] Appendix D provides a complete summary of the 12-month frequency-of-use models.

- **Predicted Means:** The predicted mean that resulted from the 12-month frequency-of-use model was a logit of the proportion of the year used. This logit was back-transformed into a proportion for use as the variable from which the neighborhoods were created. This proportion could be treated as a probability, which, in turn, could be multiplied by the probability of past year use to make the predicted mean conditional on lifetime use of the drug in question. When calculating predicted means for some item nonrespondents, sometimes it was not known whether they were past year users. Hence, to make the predicted means conditional on the same recency of use, the variables were transformed to make them conditional on what was known.

6.3.5.2.3 Provisional Hot-Deck Step

For imputation sets that included both 12-month frequency and 30-day frequency—alcohol (Set 11), inhalants (13), marijuana (15), hallucinogens (17), cocaine and crack (27), and heroin (29)—it was necessary to provisionally impute the 12-month frequency-of-use variable so that it could be used as a covariate in the 30-day frequency-of-use imputations.

The logical constraints involved the interview date, month of first use, birthday, recency of use, and 30-day frequency of use. The likeness constraints used in the assignment of values for 12-month frequency of use were similar to those used for recency of use. State-rank groups were again based on level of past month usage. Recipients and donors were also required to have the same recency of use (past month vs. past year but not past month), whether that recency of use was reported or imputed.[100] If no donors were available within these constraints, then they were loosened in the following order: (1) the delta constraint was removed, (2) donors and recipients were no longer required to be from States with similar usage levels, and (3) donors and recipients were no longer required to have the same recency of use.

6.3.5.2.4 Assignment of Provisional Imputed Values

For all drug use measures except 12-month frequency, the observed value of interest was donated directly to the recipient. However, because donors and recipients could potentially have had a different maximum possible number of days in the year that they could have used a substance, the observed proportion of the total period was donated rather than the observed 12-month frequency. In the assignment step, the donor's proportion of the total period was multiplied by the recipient's maximum possible number of days in the year on which he or she could have used the substance in order to arrive at a 12-month frequency-of-use value for the recipient.

---

[99] As with the recency-of-use models, for a few cases, the State-rank variable could not be included in the model. Usually, but not always, the age group/drug combination that had problems was the same for recency of use and 12-month frequency of use.

[100] Because all respondents in the 12-month frequency-of-use imputation were past year users by definition, item nonrespondents who were past month users required donors who were past month users, and item nonrespondents who were past year but not past month users required donors who matched that specific recency category.

Occasionally, more than one substance was associated with a single predicted mean. However, for the provisionally imputed values, only the parent drug was used as a covariate in later models. Therefore, multivariate assignments were not needed in the provisional hot-deck step, but they did occur in the final hot-deck step for recency and frequency. For example, the recency and frequency variables for cocaine and crack formed a single imputation set (27). Although 12-month frequency questions were asked for both cocaine and crack, only the 12-month frequency for cocaine was modeled, and only the 12-month frequency for cocaine was used as a covariate in the subsequent PRD model (30-day cocaine frequency). This means that there was no need to impute a provisional value for 12-month frequency for crack.

For pain relievers, tranquilizers, stimulants, and sedatives, no provisional assignment of imputed values was necessary, because these drugs did not include a measure for 30-day frequency (Table 6.5).

6.3.5.3 30-Day Frequency of Use

6.3.5.3.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the 30-day drug frequency-of-use variables was the preliminary analysis weight, PANALWT. As with cigarettes, the file was first reduced to the domain of past month users, as defined by the provisional recency variable. Next, item respondents and nonrespondents were defined according to the same criterion used for the recency and 12-month frequency imputations. To be an item respondent, the individual had to have provided valid responses to all applicable measures for the drug of interest. The item response propensity adjustment was then computed so that the respondents' weights accurately represented all past month users of the drug. In contrast with the RP model for cigarettes, the provisional 12-month frequency was included as a covariate for those drug modules that asked the respondent to report this measure (Table 6.5). In addition, recencies of use for cigarettes, smokeless tobacco, cigars, pipes, alcohol, marijuana, cocaine, crack, heroin, hallucinogens, inhalants, pain relievers, tranquilizers, stimulants, and sedatives were included if available.[101] See Appendix D for details of the covariates used in the RP model for these variables.

6.3.5.3.2 Prediction Step

As with cigarettes, the empirical distribution for 30-day frequency of use for chewing tobacco and snuff was a mixture distribution, with a positively skewed distribution from 1 to 29 and a spike at 30. For both chewing tobacco and snuff, two models were fit. The first model determined daily versus nondaily use among past month users and the second model was used for nondaily past month users. For this second model, the response variable being modeled was the logit of the proportion of the period (30 days) during which the respondent used the substance. All other drugs that included a measure for 30-day frequency used a single model for all past month users.

[101] If the recency of use for a particular drug was not yet defined, the lifetime indication of use was used instead. The recency of use of the drug being modeled was not used, because all respondents in the model were past month users.

6.3.5.3.3 Provisional Hot-Deck Step

The only drug for which provisional 30-day frequency values were required was alcohol, because provisional 30-day frequencies were required to calculate 30-day binge drinking provisional values. Neighborhoods were created for each alcohol item nonrespondent using univariate matching. The predicted means used to create the neighborhoods were given by the product of the predicted proportion of the month used (conditioned on past month use) and the probability of past month use given lifetime use (taken from the recency-of-use models).

A logical constraint required that the donated 30-day frequency was less than or equal to the respondent's preexisting 12-month frequency and greater than or equal to the respondent's preexisting 30-day binge drinking frequency. The likeness constraints were similar to those used in the provisional hot-deck step for 12-month frequency and were loosened in the following order: (1) the delta constraint was removed, and (2) donors and recipients were no longer required to be from States with similar usage levels.

6.3.5.3.4 Assignment of Provisional Imputed Values

Although more than one substance was occasionally associated with a single predicted mean, the provisionally imputed 30-day frequencies were required only if they were needed as covariates in a subsequent model. Of the substances within the multivariate set of recency of use and frequencies of use, only alcohol contained a measure (30-day binge drinking frequency) that was lower in the sequence than 30-day frequency of use. Because alcohol is not a parent/child drug, no multivariate assignments were required for provisionally imputed 30-day frequency.

6.3.5.4 30-Day Binge Drinking Frequency

For alcohol, an additional variable was defined that measured level of usage. In particular, the variable DR5DAY measured the binge drinking frequency or the number of days in the past month during which the respondent had five or more drinks. The imputation of the 30-day binge drinking frequency was similar to the imputation of 30-day frequency of alcohol use.

6.3.5.4.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the 30-day binge drinking frequency variable was the preliminary analysis weight, PANALWT. The response propensity model was built using all past month users of alcohol, whether they were binge drinkers or not. Item respondents for alcohol were defined across recency, 12-month frequency, 30-day frequency, and the 30-day binge drinking frequency measures. Therefore, the weight adjustment used in the modeling of the 30-day binge drinking frequency was the same as was used for the 30-day frequency model. See Tables D.16, D.17, and D.18 in Appendix D for details of the covariates used in the RP model for this variable.

6.3.5.4.2 Prediction Step

The response variable of interest in the 30-day binge drinking frequency model, prior to a normalizing transformation, was the proportion of the days in a month (30) on which a respondent had five or more drinks. The range of values for the proportion was from 0 to 1,

inclusive. Hence, to model 30-day binge drinking frequency of use, the following empirical logit transformation was computed for all respondents:

$$\log\left[\left(Y_i + 0.5\right)/\left(N - Y_i + 0.5\right)\right],$$

where Y_i was the observed 30-day binge drinking frequency for respondent i and N was 30, the total number of days in the month that the respondent could have binge drunk. This transformation was nearly equivalent to the standard logit transformation:

$$Y_i^* = \log\left[P_i/\left(1 - P_i\right)\right],$$

where P_i was defined as the proportion of days in the past month during which respondent i had five or more drinks. The standard logit transformation was not used, because it was not defined for daily binge drinkers nor was it defined for nonbinge drinkers among past month users.[102] Using the adjusted weights, a linear univariate regression model was then fitted for the log-transformed variable Y_i within each age group.

The predicted means from this model were used solely in the multivariate predictive mean vector used in the final hot-deck step. No provisional imputed values were determined.

6.3.5.5 Final Hot-Deck Step

The same principles that applied to the final hot-deck step for cigarettes applied to other drugs. However, for substances with child drugs and substances with both 12-month and 30-day frequencies, the logical constraints were considerably more complicated, the predictive mean vectors were larger, and the number of missingness patterns was greater. Appendix E provides detailed information on these hot-deck steps, and Table 6.8 provides a listing of the full predictive mean vector as applied to all final hot-deck programs for recency and frequency.

The construction of the predictive mean vectors for certain drugs was often complex. The main reason for the complexity is that recency and frequency models were not fit for all child drugs. In fact, the predicted means from the models for the parent drug were often used as surrogates for the child drug predicted means to reduce the number of models that needed to be fit and to avoid convergence problems with small sample sizes for some of the rarer child drugs. For example, if the individual requiring imputation is a past year user of cocaine but he or she has a missing crack recency, then the predictive mean vector includes the probability of past month cocaine use, given that the individual is a past year user of cocaine. When constructing the predictive mean vectors, the following general principles were followed:

1. If both the parent drug recency and the child drug recency(ies) were missing, condition on the general recency category of the parent drug.

[102] If the respondent was a daily binge drinker of alcohol, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[(N + 0.5)/0.5]$, where Y was the observed 30-day binge drinking frequency and N was the total number of days that the respondent could have used (usually 30). If the proportion was zero, then $\log[(Y + 0.5)/(N - Y + 0.5)] \approx \log[0.5/(N + 0.5)]$. See Cox and Snell (1989) for a discussion of the empirical logit transformation.

2. For smokeless tobacco, if both the chewing tobacco recency and the snuff recency were missing, condition on the union of the two sets of possible specific recency categories. For example, if chewing tobacco recency was "past year" and snuff recency was "past 3 years but not past month," condition on use in the past 3 years.
3. Condition all elements of the predictive mean vector on the same general recency level.

Table 6.8 Elements of Full Predictive Mean Vector

Drug Use Measure and Category of Interest	Predicted Mean	Substance
Recency of Use, Past Month Use[1]	*P*(past month user | lifetime user)	All substances
Recency of Use, Past Year but Not Past Month Use[1]	*P*(past year but not past month user | lifetime user)	All substances except pipes
Recency of Use, Past 3 Years but Not Past Year Use[1]	*P*(past 3 years but not past year user | lifetime user)	Tobacco products[2] only
12-Month Frequency of Use	*P*(use on a given day in the year | past year user) × *P*(past year user | lifetime user)[3]	All substances except tobacco
30-Day Frequency of Use for Alcohol and Substances with Few Daily Users[4]	*P*(use on a given day in the month | past month user) × *P*(past month user | lifetime user)[5]	All substances except cigarettes, chew,[6] snuff, pipes, and pills[7]
30-Day Frequency of Use for Substances with Many Daily Users (excluding Alcohol)	*P*(use on a given day in the month | past month user, not a daily user) × *P*(not a daily user | lifetime user) × *P*(past month user | lifetime user)[5]	Cigarettes, chewing tobacco, snuff
Daily User over Past 30 Days	*P*(daily user | past month user) × *P*(past month user | lifetime user)[5]	Cigarettes, chewing tobacco, snuff
30-Day Binge Drinking Frequency	*P*(drank 5 or more drinks on a given day in the past month | past month user) × *P*(past month user | lifetime user)[5]	Alcohol only

[1] The final category for recency (lifetime but not past year or lifetime but not past 3 years) was not needed in the predictive mean vector, because the multinomial probabilities summed to 1, and this probability was determined by the other probabilities.
[2] "Tobacco products" includes cigarettes, cigars, chewing tobacco, and snuff.
[3] Interpreting the proportion of the year used as a probability of use on a given day in the year assumed that the probability of use on each day in the year was equal. However, this was not true. The violation of this assumption did not seriously affect the ability to find a reasonable variable to use for finding a neighborhood, and it did allow the predicted mean to be made conditional on what was known.
[4] Alcohol, with many daily users, was included in this group because the distribution did not show a severe drop-off from 30 days a month to 29 days a month, as was apparent with cigarettes, chewing tobacco, and snuff.
[5] Interpreting the proportion of the month used as a probability of use on a given day in the month assumed that the probability of use on each day in the month was equal, which was not true, in the same manner as the 12-month frequency of use (see note #3 for this table).
[6] "Chew" is short for "chewing tobacco."
[7] "Pills" includes pain relievers, tranquilizers, stimulants, and sedatives.

6.3.5.6 Final Variables

Similar to the final imputation-revised recency-of-use and 30-day frequency variables for cigarettes (IRCIGRC and IRCIGFM), the final imputation-revised recency and frequency variables for other drugs were identified with the prefix IR, followed by a five-letter identifier, where a three-letter code identified the drug and the final two letters identified the measure. In addition to the RC and FM identifiers used for cigarettes, the identifier FY was used for the 12-month frequency variable. Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.3.5.7 Recodes for Additional Analyses

Section 6.3.2.9 lists three dichotomous indicator variables that were created to indicate cigarette use in the lifetime (CIGFLAG), past year (CIGYEAR), or past month (CIGMON). Analogous variables were also created for each drug for which an imputation-revised recency was created.

Several other prevalence recodes, which covered the same three measures, were created to incorporate information from several different drugs. Table 6.9 lists these recodes and the recency variables that were used to create them. The creation of these variables was also straightforward. If the respondent was a lifetime user of any of the drugs, then the FLAG variable was set to 1; otherwise, it was set to 0. The YR and MON variables were processed in a similar manner.

Table 6.9 Prevalence Recodes Incorporating More than One Recency Variable

General Drug Category	Variable Names	Source Recency Variables
Tobacco	TOBFLAG, TOBYR, TOBMON	Cigarettes, smokeless tobacco, cigars, pipes
Psychotherapeutics	PSYFLAG2, PSYYR2, PSYMON2[1]	Pain relievers, tranquilizers, stimulants, sedatives
Illicit Drugs Other than Marijuana	IEMFLAG, IEMYR, IEMMON	Psychotherapeutics, plus inhalants, hallucinogens, cocaine, and heroin
Illicit Drugs, but Only Marijuana	MJOFLAG, MJOYR, MJOMON	Same as MRJFLAG, MRJYR, and MRJMON, except set to 0 if the corresponding IEM variable is equal to 1
Illicit Drugs	SUMFLAG, SUMYR, SUMMON	Illicit drugs other than marijuana, plus marijuana

[1] These variable names include a suffix of "2" to distinguish them from earlier versions of psychotherapeutics recodes.

6.3.6 "Other" Drugs Age at First Use

The age-at-first-use variables for imputation sets 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, and 30 in Table 6.5 (smokeless tobacco, cigars, alcohol, inhalants, marijuana, hallucinogens, pain relievers, tranquilizers, stimulants, cocaine, and heroin, respectively) were imputed in a manner similar to that described for cigarettes. However, some deviations from the process described for cigarettes applied to these "other" drugs age at first use as described below.

6.3.6.1 Response Propensity Step

The input to the weight adjustment model in the response propensity step for the age-at-first-use variables for other drugs was the preliminary analysis weight, PANALWT. The RP step for age at first use for other drugs was very similar to the RP step for cigarettes age at first use (Section 6.3.3.1). For substances that included child drugs, a response had to be provided for each drug for an individual to be considered a respondent for imputation purposes. Appendix D provides a complete list of covariates used in each model to properly adjust the weights.

6.3.6.2 Prediction Step

The PRD step for age at first use for other drugs was also very similar to the analogous step for cigarettes (Section 6.3.3.2). For substances with child drugs, only the parent drug was modeled. Modified versions of the 12-month frequency of use (where applicable) and AFU of previously imputed drugs were used as covariates and were defined as follows:

$\mathbf{new12_i}$	=	0	if respondent *i* did **not** use the drug of interest in the past 12 months
	=	12-month frequency	if respondent *i* used the drug of interest in the past 12 months
$\mathbf{AFU_i}$	=	0	if respondent *i* is **not** a lifetime drug user of the drug of interest
	=	age at first use	if respondent *i* is a lifetime drug user of the drug of interest

6.3.6.3 Hot-Deck Step

For smokeless tobacco (chewing tobacco and snuff), cocaine (crack), hallucinogens (LSD, PCP, and Ecstasy), pain relievers (OxyContin®), and stimulants (methamphetamine), more than one age-at-first-use variable was associated with a single predicted age at first use. This led to a multivariate assignment of the imputed values.

- One model for smokeless tobacco was fitted rather than individual models for chewing tobacco and snuff. The item nonrespondent received values from the donor for both chewing tobacco (if missing) and snuff (if missing), and the age at first use for smokeless tobacco was obtained by taking the minimum age at first use from chewing tobacco and snuff. Respondents were never asked directly for their age at first use for smokeless tobacco.
- For other substances with child drugs, respondents were asked for the age at first use for the parent drug and were also asked for their age at first use for each child drug. This often led to complex and numerous logical constraints. These constraints used not only parent and child ages at first use but also imputation-revised recencies and frequencies.

6.3.6.4 Year-of-First-Use, Month-of-First-Use, and Day-of-First-Use Assignments

The general principles described in Section 6.3.3.4 applied to the remaining drugs with the following exceptions.

- For smokeless tobacco, the minimum of the chewing tobacco and snuff dates was used to produce the smokeless tobacco date of first use.
- For all child drugs (daily cigarettes, LSD, PCP, ecstasy, OxyContin®, methamphetamine, and crack), the corresponding parent drug's date of first use was assigned first. Then, in the setting of the *earliest possible date* for the child drug, the parent drug's date of first use was used as an additional bound. This was done to ensure that the child drug's date of first use was never earlier than the parent drug's date of first use.
- For all parent drugs whose child drugs had recency and frequency information (hallucinogens, pain relievers, stimulants, and cocaine), the child drug recency and frequency information was used to bound the *latest possible date* for the date of first use. For example, respondents with LSD recency = 3 (i.e., lifetime but not past year user of LSD) could not have first used hallucinogens within the past year, regardless of the hallucinogens recency value. The bound created using the child drug recency and frequency was calculated in exactly the same way as for the parent recency and frequency information.
- For hallucinogens, pain relievers, and stimulants, an indicator of lifetime use of drugs other than the child drugs was created (Table 6.1). For pain relievers and stimulants, if the respondent was not a lifetime user of the "other" drugs, then the child drug's date of first use was logically assigned to the parent drug's date of first use. The handling of the child drugs for hallucinogens was more complex, because there was more than one of them. The algorithm follows:

    1. The date of first use was assigned for overall hallucinogens.
    2. The *earliest possible date*, *latest possible date*, and the final date of first use for each child drug for which the respondent was a lifetime user were assigned.
    3. For respondents who were lifetime nonusers of other hallucinogens, it was determined which, if any, child drug could have had the same date of first use as hallucinogens. Specifically, it was determined whether the date of first use for hallucinogens was between *earliest possible date* and *latest possible date* for each child drug. If none of the child drugs were eligible to receive the hallucinogens date of first use, nothing was done. Otherwise, one of the eligible child drugs was chosen at random, and its date of first use was overwritten with the hallucinogens date.

6.3.6.5 Final Age and Date-of-First-Use Variables

The final imputation-revised date-of-first-use variables for "other" variables were named in the same manner as those for cigarettes: with the prefix IR, followed by a three-letter code

identifying the drug and the final three letters identifying the measure (AGE = age at first use, YFU = year of first use, MFU = month of first use, DFU = day of first use). Again, each IR variable was accompanied by an imputation indicator with the requisite II prefix.

6.3.6.6 Recodes for Additional Analyses

Section 6.3.5.7 discusses some prevalence recodes that incorporated information from several different drugs. Incidence recodes were also created that incorporated information from multiple drugs. These incidence recodes were created for only PSY, IEM, and SUM (Table 6.9).

The age-at-first-use recodes were simply set to the minimum of the source age-at-first-use variables, and they were named with the suffix AGE: PSYAGE2, IEMAGE, and SUMAGE. For example, PSYAGE2 = minimum of IRANLAGE, IRTRNAGE, IRSTMAGE, and IRSEDAGE.

To set the date-of-first-use variables, the earliest date of first use was found among the source variables for which the respondent was a lifetime user, and the new YFU, MFU, and DFU variables were determined using the YEAR, MONTH, and DAY functions in SAS. For example, PSYYFU2 = YEAR (minimum of dates of first use of pain relievers, tranquilizers, stimulants, and sedatives).

6.3.7 Special Section: Core-Plus-Noncore Methamphetamine and Stimulants Lifetime Use and Recency of Use (Imputation Sets 31 and 32)

New questions were added to the noncore special drugs module in the 2005 NSDUH to capture information from respondents who may have used methamphetamine but did not recognize it as a prescription drug and therefore did not report use in the core stimulants module. Additional follow-up items were added in the 2006 NSDUH to resolve inconsistencies between responses regarding methamphetamine in the core stimulants module and the noncore special drugs module. These additional methamphetamine questions asked about 12-month frequency, age at first use, and date of first use.

Findings from the methamphetamine analysis report (Ruppenkamp et al., 2007) showed that it would be important to use responses from the noncore special drugs module in order to determine the best estimate of the prevalence of methamphetamine use from the NSDUH. Therefore, after the normal imputation processing of the drug variables was complete, new imputation-revised versions of lifetime use and recency-of-use variables for both methamphetamine and stimulants were created, which incorporated responses from the noncore special drugs module as well as the core module. Estimates of nonmedical use of psychotherapeutics, stimulants, and methamphetamine in the detailed tables include data from the new methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report (Office of Applied Studies, 2008). This situation is unique for lifetime use and recency of use in that it incorporates information from the noncore section of the questionnaire. For more information on the reporting of methamphetamine prevalence in the 2012 NSDUH, see Section B.4.8 of Appendix B of the 2008 national findings (Office of Applied Studies, 2009).

A detailed description of the creation of these imputation-revised variables follows. The approach used was similar to the process used in normal processing with the following exceptions:

- The provisionally imputed values for lifetime use and recency of use for core-plus-noncore stimulants and methamphetamine were used as the final imputation-revised variables.
- A different set of edited variables was used as the base for imputation.

The use of provisionally imputed values as the final imputation-revised variables for lifetime use and recency of use was implemented beginning in 2011 to eliminate the need to reimpute variables that were not used in subsequent analyses. Before 2011, lifetime drug use models were refit for stimulants, sedatives, cocaine, crack, and heroin, and provisional imputations were performed. After these models were refit, the lifetime use indicators for all drugs were reimputed using the PMN Type 3 methodology outlined in Section 3.4.3 to incorporate the noncore methamphetamine and stimulants questions. However, the only imputation-revised lifetime use questions used in further processing were the ones for stimulants and methamphetamine. Similarly, the imputation of core-plus-noncore recency variables for stimulants and methamphetamine proceeded in the same manner as the core-only variables. This process included an RP and PRD step for recency of use, followed by a provisional imputation. A response-propensity adjustment and prediction model were then fit for 12-month frequency of use, and the final core-plus-noncore recency and 12-month frequency variables were imputed in a final hot-deck step that incorporated additional noncore variables as logical constraints.

The simpler imputation methods were implemented after an impact assessment was conducted using 2010 data. For the results of this analysis, see Section 5.3.7 of the 2011 imputation report (Frechtel et al., 2013).

6.3.7.1 Final Creation of Base Variables for Imputation

The edited recency-of-use variables MTHREC06 and STMREC06, created by the editing team, were used as a starting point for the final creation of the base variables for imputation. These variables are described in Section 6.2.6.2. They are similar to METHREC and STIMREC, the edited recency-of-use variables used in normal processing, except that they incorporate responses from the noncore special drugs module and the core module.

The final base variable for imputation of lifetime use of methamphetamine was called EDMTHLIFE. It was created as follows:

EDMTHLIFE =

- 1 (lifetime user), if MTHREC06 was 1, 2, 3, 8, 9, 11, 12, or 13; else
- 2 (lifetime nonuser), if MTHREC06 was 81 or 91; else
- missing.

The final base variable for imputation of lifetime use of stimulants, EDSTMLIFE, was created in a similar manner.

The final base variable for imputation of recency of use of methamphetamine was called EDMTHREC. It was created as follows:

EDMTHREC =

- 1 (past month user), if MTHREC06 was 1, or if MTHREC06 was 11 and METHREC was not equal to 11; else
- 2 (past year but not past month user), if MTHREC06 was 2, or if MTHREC06 was 12 and METHREC was not equal to 12; else
- 3 (lifetime but not past year user), if MTHREC06 was 3 or 13; else
- MTHREC06.

Note that respondents who responded to the noncore recency question (most of those with MTHREC06 values of 11, 12, and 13) were treated identically to respondents who gave the same response to the core recency question (those with MTHREC06 values of 1, 2, and 3). This was done based on the decision to treat respondents to the noncore questions as item respondents eligible to be donors and therefore used to fit the models. This is an exception to the general rule that respondents with logically assigned responses were treated as item *non*respondents.

The final base variable for imputation of recency of use of stimulants, called EDSTMREC, was created in a similar manner.

6.3.7.2 Reimputation of Lifetime Use Indicators (Imputation Set 31)

Using EDMTHLIFE and EDSTMLIFE, the processing of the lifetime use indicators proceeded, as described in Section 6.3.1. The set of item respondents did not change between the original imputation of the lifetime indicators and the reimputation of the lifetime indicators. Therefore, it was not necessary to readjust the weights for item nonresponse. As shown in Table 6.5, the stimulants lifetime drug use indicator was modeled toward the end of the hierarchy. Rather than reimputing stimulants and all variables that came after it, lifetime models were refit for stimulants and methamphetamine only, and missing values were imputed in one univariate hot-deck step.

6.3.7.3 Reimputation of Recency of Use

Using EDMTHREC and EDSTMREC instead of METHREC and STIMREC, the processing of the recency data proceeded, as described previously.[103] Final recency-of-use variables for methamphetamine and stimulants were created.

[103] Other core-plus-noncore edited variables also were used in these reimputation steps, in logical constraints of hot-deck steps.

6.3.7.4 Recodes for Additional Analyses

In the manner described in Section 6.3.5.7, some prevalence recodes were created that incorporated information from the noncore special drugs module. The core-plus-noncore methamphetamine recodes were CPNMTHFG, CPNMTHYR, and CPNMTHMN. The core-plus-noncore stimulants recodes were CPNSTMFG, CPNSTMYR, and CPNSTMMN. The core-plus-noncore psychotherapeutic recodes were CPNPSYFG, CPNPSYYR, and CPNPSYMN. No core-plus-noncore versions of the IEM or SUM recodes described in Table 6.9 were created for use in the detailed tables, even though the prevalence estimates would likely increase slightly if the noncore methamphetamine data were incorporated.

7. Editing and Imputation for Variables from the NSDUH Noncore ACASI Modules

7.1 Introduction

The 2012 computer-assisted interviewing (CAI) instrument contained the following noncore self-administered modules, which used audio computer-assisted self-interviewing (ACASI; see Chapter 1):

- special drugs,
- risk/availability,
- blunts,
- substance dependence and abuse,
- special topics,
- marijuana purchases,
- prior substance use,
- substance treatment,
- health care,
- adult mental health service utilization (administered only to adults aged 18 or older),
- social environment (administered only to adults),
- parenting experiences (administered only to parent/legal guardian in dwelling units where a 12 to 17 year old also was selected for an interview),
- youth experiences (administered only to youths aged 12 to 17),
- mental health (administered only to adults),
- adult depression and adolescent depression (the former administered to adults and the latter administered to youths),
- youth mental health service utilization (administered only to youths), and
- consumption of alcohol.

This chapter describes the content of the individual noncore ACASI modules. Except for the substance dependence and abuse module, no imputation for missing data was performed for variables in these modules. Therefore, the focus of the discussion in this chapter is on processing of the edited variables for these modules, along with any specific edits that were relevant to the data in a given module. The structure of this chapter also differs from the structure of other chapters beginning with Chapter 4 because only one set of noncore ACASI variables was imputed. In this chapter, subheadings in Section 7.4 focus on the specific noncore ACASI

modules. Section 7.4.4 for the substance dependence and abuse module also discusses imputation procedures for the cigarette (i.e., nicotine) dependence items in this module.

7.2 Editing of Noncore ACASI Data Because of Patterned Responses

Two types of edits could be made to noncore ACASI data because of patterned responses in the interview. The first type involved edits because of patterned responses in corresponding variables from core modules (see Section 2.3.2 in Chapter 2). When a case was retained as a final respondent but with variables from one or more core modules being assigned the appropriate codes for bad data (Section 2.4.3), codes for bad data also were assigned to the corresponding variables in noncore modules. This edit was implemented regardless of whether the corresponding noncore variables had patterned responses. For example, if a respondent's variables in the pain relievers module were assigned codes for bad data because of patterned responses in that module and the respondent was routed to questions pertaining to pain relievers in the substance dependence and abuse module, codes for bad data were assigned to the pain relievers data in the substance dependence and abuse module.

The second type of edit involved edits because of other occurrences of patterned responses specifically in the noncore ACASI data. These edits were implemented for a small number of cases in 2012 (i.e., fewer than 20). Responses were replaced with bad data codes when these patterns occurred. In particular, answers were replaced with bad data codes if respondents keyed answers of "1" (or multiple-digit responses of "1," where applicable) to every question they were asked within a module. In the youth experiences module, adolescents who keyed answers of "2" wherever possible also had their answers replaced with bad data codes. For respondents to key all possible items as "2" in the youth experiences module would mean that they keyed this response to 42 items in a row, including that they had engaged in all "problem" activities (such as getting into fights, carrying a handgun, or selling illegal drugs) exactly "1 or 2 times" in the past year. These respondents who keyed "2" wherever possible also would have participated in exactly two school-based activities, two community-based activities, two church- or faith-based activities, and two other kinds of activities in the past year.

7.3 "Lockout" Following Completion of ACASI Modules

Since 2003, once the ACASI portion of the interview had been completed, interviewers and respondents are locked out of the ACASI section of the interview to provide additional protection of the privacy of respondents' answers. Based on answers that respondents gave in the interviewer-administered questions following the ACASI section, however, interviewers could go back and change information in core demographics questions that were in the interviewer-administered section that preceded the ACASI questions. For example, interviewers could discover when they were completing the household roster section of the interview that a female respondent's gender had been incorrectly keyed as male, or vice versa (see Section 4.2.4). The interviewer was allowed to change the information about the respondent's gender at the beginning of the interview.

If interviewers went back and changed answers to core demographic questions, however, the CAI program did not reroute respondents back into the ACASI section to answer questions that previously had been skipped. The CAI program also did not update the answers in the

ACASI section to remove data from questions that logically should have been skipped based on any new information that the interviewers entered. This "lockout" feature was particularly relevant to editing of data in the modules for health care (Section 7.4.9), parenting experiences (Section 7.4.12), and consumption of alcohol (Section 7.4.17).

7.4 Editing and Imputation Procedures for Specific Noncore ACASI Modules

7.4.1 Special Drugs Module

The special drugs module asked about the smoking and sniffing of heroin; use of heroin, methamphetamine, other stimulants, cocaine, or other drugs with a needle for nonmedical reasons; general needle use behaviors (e.g., needle sharing); and where respondents got the last needle that they used. Questions also have been included in the module since 2006 for additional drugs that did not have specific questions in any of the core drug modules.

As noted in Sections 6.1.3 and 6.2.6 in Chapter 6, new questions were added to the special drugs module in 2005 and were updated in 2006 to capture information about methamphetamine use from respondents who did not report methamphetamine use in the core stimulants module (e.g., if they may not have recognized it as a stimulant in the context of questions about prescription stimulants). Details about the logic for asking these follow-up questions about methamphetamine use are discussed in Section 6.2.6.

Additional questions have been included in the special drugs module since 2007 that ask about first use of methamphetamine and frequency of use in the past 12 months. However, these items on first use of methamphetamine and 12-month frequency were not used in editing the methamphetamine recency variable MTHAREC that was described in Section 6.2.6. Thus, MTHAREC was created in a manner comparable with how it was created since 2006 to maintain data on trends in methamphetamine use. However, inconsistencies could exist between MTHAREC and these new variables, such as if respondents reported that they initiated methamphetamine use in the past 12 months but they reported last using methamphetamine more than 12 months ago.

7.4.1.1 Noncore Methamphetamine Initiation and 12-Month Frequency

A key aspect of processing the noncore variables about first use of methamphetamine and frequency of use in the past 12 months involved assignment of appropriate codes of 91 (or 991, etc.) if respondents never used methamphetamine (Section 2.4.2). In addition, codes of 93 (or 993) were assigned to the 12-month frequency variables if respondents last used methamphetamine more than 12 months ago.

Respondents who confirmed methamphetamine use in the consistency check questions (SD17A1 or SD17ALT) were asked to report their age when they first used methamphetamine (edited variable MTHAAGE). Respondents who first used within 1 year of their current age were asked to report the year and month when they first used methamphetamine (edited variables MTHAYFU and MTHAMFU, respectively). The CAI logic triggered consistency checks if respondents reported a year and month of first use that was inconsistent with their reported age at

first use. Thus, these items were analogous to those that asked about first use of drugs in the core modules (Section 6.2.5). Consequently, the same edits were applied to these methamphetamine variables as were applied to the corresponding questions in the core modules that were described in Section 6.2.5.

Additionally, respondents who reported using methamphetamine in the past 12 months in questions SD17B or SD18B and who did not report that an earlier answer that they never used methamphetamine was correct (as reported in the core stimulants module) were asked to report on their frequency of use of methamphetamine in the past 12 months. As in the core drug modules, respondents could report their methamphetamine frequency of use in one of three ways:

- use on an average number of days per week in the past 12 months,
- use on an average number of days per month in the past 12 months, or
- the total number of days they used in the past 12 months.

As for the 12-month frequency variables in the core modules (Section 6.2.4), a single composite edited variable MTHAYTOT was created to translate these answers to a total number of days that respondents used methamphetamine in the past 12 months. In the same manner as for 12-month frequency variables in the core drug modules, the value in MTHAYTOT was prorated as needed for respondents who initiated use in the past 12 months and therefore were not users over the entire 12-month period (see Table B.5 in Appendix B).

In addition, the CAI logic for asking about first use of methamphetamine and the 12-month frequency of use differed slightly with respect to the consistency checks SD17A1 and SD17ALT that were triggered when respondents reported in the core stimulants module that they never used methamphetamine but reported use in the special drugs module. Specifically, the only respondents who were eligible to be asked about first use of methamphetamine in the special drugs module were those who specifically reported in the core stimulants module that they never used methamphetamine and then explicitly confirmed their methamphetamine use in SD17A1 or SD17ALT. However, respondents who reported last using methamphetamine in the past 12 months were eligible to be asked about their 12-month frequency, as long as they had not reported in SD17A1 or SD17ALT that an earlier report of having never used stimulants was correct (see Section 6.2.6); this included situations in which SD17A1 and SD17ALT had been skipped because respondents reported in the core stimulants module that they did not know whether they ever used methamphetamine (i.e., and therefore had not previously denied ever using the drug).

7.4.1.2 Heroin Use and Needle Use Variables

An important aspect of the processing of variables in this module pertaining to the smoking of heroin, the sniffing of heroin, or needle use consisted of assigning codes of 91, 93, or 99 (see Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, respondents who never used heroin, stimulants other than methamphetamine, or cocaine were not asked questions in the special drugs module that pertained to these drugs. Similarly, respondents who did not indicate use of methamphetamine in the core stimulants

module and who continued to indicate in the follow-up questions in the special drugs module that they never used methamphetamine did not need to be asked further questions pertaining to the use of methamphetamine with a needle. In addition, respondents who indicated that they never used heroin, methamphetamine (in both the core stimulants module and in the noncore special drugs module), stimulants other than methamphetamine, cocaine, or any other drug with a needle for nonmedical reasons did not need to be asked questions about general needle use behaviors or the source of the last needle they used. Table B.13 in Appendix B describes specific edits that were implemented for these variables when the corresponding items were skipped.

Respondents who reported in the core heroin module that they were lifetime heroin users but who answered "no" to all questions about smoking heroin (question SD01), sniffing heroin (question SD03), or using it with a needle (question SD08) were asked a follow-up question SDHEUSE to determine how these respondents administered the heroin they had reported using. SDHEUSE was an "enter all that apply" type of question that allowed respondents to report multiple ways that they used heroin. SDHEUSE included response options for smoking heroin, sniffing heroin, using heroin with a needle, or use of heroin "some other way." Respondents who reported using heroin "some other way" were asked to specify in question SDHEUSE2 what this "other" mode of heroin administration was. If respondents specified using heroin in a way that corresponded to one of the ways that they had been asked about, it was logically inferred that the response in the edited SDHEUSE question should have been chosen. For example, if respondents did not report in SDHEUSE that they smoked heroin but they specified this as "some other way," the edited variable HEOTSMK was assigned a code of 3 (Response entered LOGICALLY ASSIGNED).

In turn, HEOTSMK, HEOTSNF, and HEOTNDL were used to edit the corresponding lifetime heroin variables HERSMOKE, HERSNIFF, and HERNEEDL, respectively. For example, if HEOTSMK indicated that the respondent had smoked heroin (HEOTSMK = 1 or 3), HERSMOKE was edited to indicate that the respondent was logically inferred to have smoked heroin at least once in his or her lifetime.

In addition, respondents were routed to follow-up recency questions for smoking, sniffing, or using heroin with a needle if they reported using heroin in any of these ways in question SDHEUSE. Information from these follow-up questions was used in the creation of the heroin smoking, sniffing, or needle recency variables HRSMKREC, HRSNFREC, and HRNDLREC. However, if respondents initially did not report using heroin in these ways in SDHEUSE, they were skipped out of these follow-up recency questions for smoking, sniffing, or using heroin with a needle. Therefore, if respondents' only indication of smoking, sniffing, or using heroin with a needle came from the "OTHER, Specify" response associated with SDHEUSE, the corresponding variables HRSMKREC, HRSNFREC, or HRNDLREC were set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED).

Table B.14 describes specific edits other than those pertaining to skip logic that were implemented in the special drugs module for heroin use and use of drugs with a needle.[104] For

[104] Issues related to the editing of the needle use variables for methamphetamine and stimulants are discussed in Table B.11 in the context of other issues for editing the noncore special drugs variables that are predecessors to the edited core-plus-noncore recency variables for these substances.

example, lifetime users of heroin could report that they smoked heroin at least once but not indicate when they last smoked it. The general edit was to assign a nonspecific value to the edited recency variable (i.e., HRSMKREC) to indicate that the respondent smoked heroin at some point in his or her lifetime. In some special situations, however, it was possible to infer that respondents could not have smoked heroin in the past 12 months. In these situations, respondents reported last using heroin more than 12 months ago, and there were no responses for other heroin-related questions in the special drugs module to indicate that these respondents had used heroin in the past 12 months. Since 2001, respondents have been asked in question SD05 (edited variable OTDGNEDL) whether they ever used a needle to inject "some other drug" with a needle (if respondents previously reported lifetime use of heroin, methamphetamine, other stimulants, or cocaine with a needle) or "any drug" with a needle (if respondents had not previously indicated use of any of the above drugs with a needle). If question SD05 was answered as "yes" (OTDGNEDL = 1), respondents then were asked to specify what (other) drug(s) they used with a needle. Respondents could specify up to five drugs that they had injected (edited variables OTDGNDLA through OTDGNDLE).

Consequently, it was possible for respondents to have reported in a core drug module that they never used a particular drug that was covered in the special drugs module but then specify use of that drug with a needle in OTDGNDLA through OTDGNDLE. For example, respondents could indicate in the core heroin module that they never used heroin but then specify lifetime injection of heroin in OTDGNDLA through OTDGNDLE. In this situation, no editing was done to the core drug data. However, these respondents were logically inferred in the relevant special drugs variables to be users of that particular drug at some point in the lifetime. Thus, for example, if respondents reported in the core heroin module that they never used heroin, but then they specified heroin as "some other drug" that they used with a needle, the edited lifetime heroin needle use variable HERNEEDL was assigned a code of 3 (Yes LOGICALLY ASSIGNED) and the heroin needle recency variable HRNDLREC was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED).

Respondents also could report in the special drugs module that they used a needle to inject a drug for nonmedical reasons, even though they previously reported that they never used marijuana, cocaine, heroin, hallucinogens, inhalants, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives. Beginning in 2001, however, respondents were asked to specify what "other" drug(s) they had injected. Thus, it was possible to identify respondents who corroborated their report of lifetime injection drug use (e.g., if injection of anabolic steroids was reported) despite having previously reported nonuse of all drugs covered in the core modules. Similarly, it was possible from "OTHER, Specify" data on other drugs that respondents had injected to identify those whose needle use probably was limited to legitimate, medical uses (e.g., injection of antibiotics). Therefore, we logically inferred that respondents had never used needles for nonmedical purposes if (1) they were lifetime nonusers of all drugs covered in the core; (2) they indicated that they never engaged in behaviors that would be indicative of nonmedical needle use, such as needle sharing, use of bleach to clean needles, or reusing of needles; and (3) all of the "other" drugs they reported using with a needle typically were not drugs of abuse. In this situation, the variable OTDGNEDL corresponding to question SD05 was set to a value of 4 (No LOGICALLY ASSIGNED). Any data in the general needle use variables GNNDREUS through GNNDGET were replaced with codes of 89 (LEGITIMATE SKIP Logically assigned) to indicate that respondents logically should have

skipped these items because they appeared to be reporting about legitimate use of drugs with a needle.

7.4.1.3 Recoded Needle Use Variables

Recoded needle use variables (STNEELDR, STNLRECR, and NEDLRECR) were created from variables pertaining to use of methamphetamine, other stimulants, heroin, or cocaine with a needle. STNEEDLR and STNLRECR were analogous to the variables STNEEDLE and STNDLREC, respectively, that existed in 1999 and 2000. Similarly, NEDLRECR was analogous to the recoded needle recency variable NEDLRECC that existed in 1999 and 2000. These variables did not take into account data from the methamphetamine variables that have been present in the special drugs module since 2005. Consequently, STNEEDLR, STNLRECR, and NEDLRECR were designed to be comparable with data in these variables prior to 2005 for use in analysis of trends in needle use.

STNEEDLR was created from the lifetime methamphetamine needle use variable MTHNEEDL and the lifetime other stimulant needle use variable OSTNEEDL. The logic for creating STNEEDLR is documented below.

- If respondents reported use of methamphetamine or other stimulants with a needle, STNEEDLR was coded as 1 ("yes").
- If respondents were users of methamphetamine or other stimulants but reported never injecting stimulants with a needle, STNEEDLR was coded as 2 ("no").
- If respondents had never used stimulants, STNEEDLR was coded as 81 or 91 (depending on the value in the core stimulant recency variable STIMREC). Missing data from MTHNEEDL or OSTNEEDL were carried over to STNEEDLR.

STNLRECR was derived from the needle recency variables MTNDLREC (most recent use of methamphetamine with a needle) and OSTNLREC (most recent use of other stimulants with a needle). The logic for creating STNLRECR is documented below.

- If respondents had never used stimulants, STNLRECR was assigned a code of 81 or 91, depending on the value in STIMREC.
- Similarly, if STNEEDLR was coded as 2 ("no"), STNLRECR was coded as 93 (used stimulants but never with a needle).
- If respondents had injected methamphetamine or some other stimulant, the general principle in assigning a value to STNLRECR was to pick the most recent use that the respondent reported. For example, if respondents indicated in either MTNDLREC or OSTNLREC that they used methamphetamine or other stimulants with a needle within the past 30 days, then STNLRECR indicated use of stimulants with a needle in the past 30 days.
- However, if respondents reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months but all that was known was that they used other stimulants with a needle at some point in their lifetime, it still could be inferred that they had used a needle to inject any stimulant at some point in the past

12 months (potentially including the past 30 days). The nonspecific value for past year use was assigned (i.e., a code of 8) because the respondents could have been more recent users of other stimulants with a needle.

- Similarly, if respondents indicated use of one of these stimulants (i.e., methamphetamine or other stimulants) with a needle in a definite period more than 30 days ago but the respondents did not know or refused to indicate whether they had ever used the other type of stimulant with a needle, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to STNLRECR because the respondents may have been more recent users of stimulants with a needle than what they had reported.

The needle recency variable NEDLRECR was created from the variables HRNDLREC (most recent use of heroin with a needle), CONDLREC (most recent use of cocaine with a needle), and STNLRECR (most recent use of any stimulant with a needle, as noted above). The logic for creating NEDLRECR is documented below.

- If respondents never used a needle to inject any of these drugs nonmedically (including situations in which respondents never used cocaine, heroin, or stimulants), NEDLRECR indicated that the respondents had never used cocaine, heroin, or stimulants with a needle.
- If respondents reported using one or more of these drugs with a needle, the general principle in assigning a value to NEDLRECR was to identify the most recent use reported by the respondents. In particular, if respondents reported using one or more of these drugs with a needle in the past 30 days, it could be determined unambiguously that the respondents were past month needle users.
- In other situations, however, if one or more of the cocaine, heroin, or stimulant needle recency variables indicated nonspecific use at some point in the respondents' lifetime, NEDLRECR was assigned a value to indicate nonspecific past year or lifetime use. For example, if respondents reported last using cocaine with a needle more than 30 days ago but within the past 12 months, yet all that was known was that they used heroin with a needle at some point in their lifetime, it still could be inferred that they had used some drug with a needle in the past 12 months. The nonspecific value for past year use was assigned (i.e., a code of 8) because the respondents could have been more recent users of heroin with a needle.
- Similarly, if respondents indicated use of one of these drugs with a needle more than 12 months ago and they did not know or refused to indicate when they last used one of the other drugs with a needle, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to NEDLRECR because the respondents may have been more recent needle users than what they had reported elsewhere.

In addition, recoded needle use variables (STNEDL06, STNLRC06, and NDLREC06) were created to incorporate the methamphetamine needle use questions that were added to the module in 2005 and the follow-up items on methamphetamine use that were added to the module in 2006 (and retained since then). These new needle use variables were based on the variables STNEELDR, STNLRECR, and NEDLRECR (documented above). Consequently, data in

STNEDL06, STNLRC06, and NDLREC06 may not be comparable with data in STNEEDLR, STNLRECR, and NEDLRECR from prior survey years.

The recoded lifetime stimulant needle variable STNEDL06 was derived from the variable STNEEDLR. STNEDL06 was updated to reflect data from the new variable MTHANEDL (ever used a needle to inject methamphetamine), after MTHANEDL had been edited (see Section 6.2.6 and Table B.11 in Appendix B). The logic for creating STNEDL06 is documented below. This logic for creating STNEDL06 was implemented only if (1) MTHEVCK had been set to bad data (see Table B.11), (2) respondents had specified use of methamphetamine as "some other drug" that they used with a needle, or (3) respondents had not previously reported methamphetamine use in the core stimulants module because they had not thought of methamphetamine as a prescription drug or for some similar reason (e.g., knowing methamphetamine by a slang term other than the examples given in the core).

- If MTHANEDL indicated use of methamphetamine with a needle and STNEEDLR did not already indicate use of any stimulant with a needle, then STNEDL06 was recoded as 1 ("yes").
- If STNEEDLR had been coded as 81 or 91 (i.e., never used stimulants) and MTHAMP indicated lifetime methamphetamine use but MTHANEDL did not indicate use with a needle, then STNEDL06 was recoded as 2 ("no"), instead of 81 or 91.
- Otherwise, if MTHANEDL had a missing value and STNEEDLR did not indicate that respondents used stimulants with a needle, STNEDL06 was assigned a missing value.

The recoded stimulant needle recency variable STNLRC06 was derived from the variable STNLRECR and was updated to reflect data from MTANDLRC (most recent use of methamphetamine with a needle), after MTANDLRC had been edited. The logic for creating STNLRC06 is documented below. The same conditions described above for implementing the logic for STNEDL06 also applied to STNLRC06.

- If respondents indicated last using methamphetamine with a needle within the past 30 days in MTANDLRC and STNLRECR did not already indicate use of stimulants with a needle in the past 30 days, then STNLRC06 was recoded to indicate use of stimulants with a needle in the past 30 days.
- If respondents indicated needle use in MTANDLRC in some period other than the past 30 days but STNLRECR indicated that respondents either had never used stimulants or had never used stimulants with a needle (i.e., based on core stimulants data and special drugs variables that existed prior to 2005), then STNLREC06 was updated with the period of most recent use that the respondents indicated in MTANDLRC.
- If respondents reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months (in MTANDLRC) but all that was known was that the respondents used other stimulants with a needle at some point in their lifetime, it still could be inferred that the respondents had used a needle to inject any stimulant at

some point in the past 12 months (potentially including the past 30 days). The nonspecific value for past year use was assigned (i.e., a code of 8) because respondents could have been more recent users of other stimulants with a needle.

- Similarly, if respondents indicated use of methamphetamine (from MTANDLRC) or other stimulants with a needle in a definite period more than 30 days ago and the respondents did not know or refused to indicate whether they had ever used the other type of stimulant, a nonspecific value of lifetime use (i.e., a code of 9) was assigned to STNLRC06 because the respondents may have been more recent users of stimulants with a needle than what they had reported.
- If MTANDLRC had a definite value indicating most recent use more than 30 days ago, STNLRECR was coded as 9 (i.e., used at some point in the lifetime), but OSTNEEDL indicated that the respondents had not used other stimulants with a needle, then the STNLRC06 was assigned the value from MTANDLRC plus a value of 10. For example, if MTANDLRC was coded as 3 (last used methamphetamine with a needle more than 12 months ago) and STNLRECR was coded as 9, then STNLREC06 was assigned a value of 13. In this situation, the code of 9 in STNLRECR would have come from the methamphetamine data that existed in the special drugs module prior to 2006, and not from data for other stimulants.
- If STNLRECR indicated that that respondents had never used stimulants (i.e., STNLRECR = 91) but MTANDLRC indicated that they had used methamphetamine but never with a needle, then STNLREC06 was recoded to a value of 93 (USED STIMULANTS BUT NEVER WITH A NEEDLE).
- If STNLRECR was coded as 91 (see above) and MTANDLRC also indicated that respondents had never used methamphetamine, then STNLREC06 continued to be coded as 91.

Because of the addition of the new methamphetamine variables, the needle recency variable NDLREC06 was derived from the variable NEDLRECR (see above). NDLREC06 was updated based on values in STNLRC06. The logic for creating NDLREC06 is documented below.

- If respondents indicated last using stimulants with a needle within the past 30 days based on STNLRC06, and NDLREC06 did not already indicate use in that period, then NDLREC06 was recoded to indicate use of cocaine, heroin, or stimulants with a needle in the past 30 days.
- If STNLRC06 indicated that respondents used stimulants with a needle at some point in the past 12 months (STNLRC06 = 8) and NEDLRECR did not indicate use in the past 30 days or NEDLRECR was not already coded as 8 (i.e., used at some point in the past 12 months), then NDLREC06 was recoded as 8.
- Similarly, if STNLRC06 indicated that respondents used stimulants with a needle at some point in their lifetime (STNLRC06 = 9) and NEDLRECR did not indicate use in the past 12 months or NEDLRECR was not already coded as 9 (i.e., use at some point in the lifetime), then NEDLREC06 was recoded as 9.

- If respondents reported using stimulants with a needle (from STNLRC06) in some definite period other than the past 30 days, the general principle in assigning a value to NDLREC06 was to identify the most recent use reported by the respondents. For example, STNLRC06 indicated that respondents last used stimulants with a needle more than 30 days ago but within the past 12 months but NEDLRECR indicated that respondents last used a needle to inject cocaine, heroin, or stimulants more than 30 days ago but within the past 12 months, then NDLREC06 was recoded to reflect the value from STNLRC06.
- If respondents were logically inferred to have last used stimulants with a needle more than 12 months ago (STNLRC06 = 13) and NEDLRECR was coded as 9 (i.e., used at some point in the lifetime) but respondents had never used cocaine or heroin with a needle, then NEDLREC06 was recoded as 13.

If STNLRC06 had a missing value and NEDLRECR indicated needle use at some point in the past 12 months (but not in the past 30 days), then NDLREC06 was recoded as 8 (i.e., used at some point in the past 12 months). Similarly, if STNLRC06 had a missing value and NEDLRECR indicated needle use but not in the past 12 months, then NDLREC06 was recoded as 9 (i.e., used at some point in the lifetime).

7.4.1.4 Additional Drug Use Variables

Questions have been included in the special drugs module since 2006 for the following drugs that did not have specific questions in any of the core drug modules:

- GHB (gamma hydroxybutyrate);
- Adderall® (a prescription stimulant) used without a prescription or for the experience or feeling it caused (i.e., "nonmedical" use);
- Ambien® (a prescription sedative) used without a prescription or for the experience or feeling it caused;
- nonprescription cough or cold medicines, also known as over-the-counter (OTC) medicines, to get high;
- ketamine (a hallucinogen);
- the hallucinogens DMT (dimethyltryptamine), AMT (alpha-methyltryptamine), and 5-MeO-DIPT (5-methoxy-diisopropyltryptamine); and
- *Salvia divinorum* (a hallucinogen).

Respondents were shown images on-screen of Adderall® and Ambien® to aid them in recalling whether they had ever used these drugs nonmedically.

If respondents reported that they ever used a particular drug mentioned above, they were asked to report when they last used that drug. In addition, respondents who reported that they last used OTC cough or cold medicines to get high in the past 12 months were asked to specify what OTC drugs they used to get high in that period. Respondents could specify up to five drugs that they used to get high. Therefore, the edited variables COLDYR1 through COLDYR5

(corresponding to the "OTHER, Specify" items SD27a through SD27e) captured up to five medications that respondents reported using to get high.

The "OTHER, Specify" variables for OTC cough and cold medicines were coded using a decimal system to preserve relationships between similar responses and to capture additional details that respondents provided. For example, there are multiple formulations of OTC cold medicines under the brand name of Robitussin® that vary in terms of the specific ingredients they contain, such as the OTC cough suppressant dextromethorphan (DXM). Therefore, if respondents specified that they used a Robitussin® product to get high in the past 12 months but the only detail they provided was that they used "Robitussin," that response was assigned a default code of 829.0 (Robitussin®, no other information). However, if respondents specifically reported that they used Robitussin® DM to get high (a product that contains DXM), that response was coded as 829.1 and was documented as Robitussin® DM. Thus, the main code of 829 indicated use of a Robitussin® product to get high in the past 12 months, and the decimal value indicated whether respondents provided additional detail about their use of specific Robitussin® products to get high.

An important aspect of the processing of these drug variables consisted of assigning codes of 91 (or 9991) to variables that had been skipped because the questions did not apply. For example, if respondents reported in question SD19 (edited variable GHB) that they never used GHB, the edited recency-of-use variable GHBREC (corresponding to question SD24) was assigned a code of 91. Similarly, if respondents reported that they last used an OTC cough or cold medicine to get high more than 12 months ago, COLDYR1 through COLDYR5 were assigned codes of 9993.0 to indicate that respondents had used OTC cough or cold medicines to get high, but not in the past 12 months.

In addition, if respondents reported that they had ever used a drug (e.g., GHB) but they did not know or refused to report when they last used it, the edited recency-of-use variable (e.g., GHBREC) was assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). If the most recent use of OTC cough or cold medicines (edited variable COLDREC, corresponding to question SD27) had been assigned a code of 9, COLDYR1 through COLDYR5 were left as blank (code of 9998.0) because some of these respondents may have used OTC cough or cold medicines to get high in the past 12 months.

Data from the "OTHER, Specify" needle use variables OTDGNDLA through OTDGNDLE and the cough and cold medicine variables COLDYR1 through COLDYR5 were used in editing these additional drug use variables, except for editing of the OTC cough and cold variables COLDMEDS (corresponding to question SD22A) and COLDREC. Specifically, OTDGNDLA through OTDGNEDLE were not used to edit COLDMEDS and COLDREC because there could be too much room for interpretation regarding what might constitute a nonprescription cough or cold medicine. In contrast, the questions for the other additional drugs in the special drugs module pertained to specific drugs.

Thus, for example, if a respondent had not reported lifetime use of ketamine but the respondent specified using ketamine with a needle in OTDGNDLA through OTDGNDLE, then the edited variable KETAMINE (corresponding to question SD23a) would be assigned a code of 3 (Yes LOGICALLY ASSIGNED). In turn, the recency-of-use variable KETAREC

(corresponding to question SD28) would be assigned a code of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). However, there were no situations in 2012 in which data from the "OTHER, Specify" needle use variables OTDGNDLA through OTDGNDLE resulted in any responses being logically inferred in these additional drug use variables. Nevertheless, this logic is in place for handling these types of patterns in future survey years.

In addition, if a respondent specified use in the past year of GHB, Adderall®, Ambien®, ketamine, DMT, AMT, 5-MeO-DIPT, or *Salvia divinorum* as a nonprescription cough or cold medicine in COLDYR1 through COLDYR5, that response was used to edit the other drug data. Using Ambien® as an example, if question SD21 pertaining to lifetime nonmedical use of Ambien® was not answered as "yes" (or had not logically been inferred as "yes" from the above edits based on OTDGNDLA through OTDGNDLE), the corresponding edited variable AMBIEN was assigned a code of 3 (Yes LOGICALLY ASSIGNED). In turn, the corresponding edited recency-of-use variable (e.g., AMBIREC, corresponding to question SD26) was assigned a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED) because COLDYR1 through COLDYR5 pertain to use in the past 12 months. If a respondent had answered question SD21 for Ambien® as "yes" and reported last using it more than 12 months ago but also specified using it in the past 12 months in COLDYR1 through COLDYR5, AMBIREC also was assigned a code of 8.

Because the questions corresponding to COLDMEDS and COLDREC asked specifically about use of OTC cough or cold medicines, respondents also would be expected to report at least one OTC medicine in COLDYR1 through COLDYR5 or to have some response of "don't know" or "refused" in the corresponding items (i.e., if respondents did not know the name of the drug or did not want to report it). Therefore, a variable called OTCFLAG was created to indicate when respondents' reports of use of OTC cough or cold medicines in the past year were not substantiated by the "OTHER, Specify" data. OTCFLAG had the values shown below when these situations occurred.

- OTCFLAG had a value of 1 (i.e., "yes") when respondents specified only the use of drugs that require a prescription in the United States[105] or drugs that are not available as OTC medicines in the United States, even if they may be available as OTC medicines elsewhere.
- OTCFLAG had a value of 2 (i.e., "no") when COLDYR1 through COLDYR5 had values only of 9985.0 (bad data) or 9998.0 (blank).

No further editing was done to the OTC variables when OTCFLAG indicated that respondents' reports of use of OTC cough or cold medicines to get high in the past year were not substantiated in COLDYR1 through COLDYR5. However, OTCFLAG would alert analysts when this issue occurred.

[105] If respondents reported the use of drugs that may require a prescription in some locations in the United States but may be available without a prescription elsewhere in the United States (e.g., cough medicines containing low doses of codeine), the medication was assumed to have been obtained without a prescription.

7.4.2 Risk/Availability Module

The risk/availability module asked about the perceived risk of harm associated with use of alcohol or specific illegal drugs, perceived ease of obtaining different illegal drugs, whether respondents were approached by someone in the past 30 days who was trying to sell an illegal drug, and general risk-taking types of behaviors. The latter included questions on the frequency with which respondents got a "kick out of doing things that are a little dangerous," how often they tried to test themselves "by doing something a little risky," and their frequency of seatbelt use.

Minimal processing of data was done to variables in this section. Specifically, the unedited variables were assigned final, mnemonic variable names (e.g., RSKPKCIG corresponding to question RK01A, which asked about the perceived risk of harm associated with smoking one or more packs of cigarettes per day). No further editing or processing was done to the variables in this module.

7.4.3 Blunts Module

Since 2004, all respondents have been asked if they ever smoked part or all of a cigar with marijuana in it, commonly referred to as a "blunt." If respondents were lifetime users of blunts, they also were asked how long it had been since they had smoked a blunt. In addition, if respondents indicated blunt use in the past 30 days, they were asked to report the number of days they smoked blunts in the past 30 days. If respondents reported use of both cigars (from the tobacco module) and blunts in the past 30 days, they also were asked if they had smoked a cigar *without* marijuana in it during the past 30 days.

In addition, the module included consistency checks in selected situations where respondents gave reports of blunt use that were inconsistent with what they had reported in the core marijuana module. Specifically, respondents could have reported earlier in the interview that they never used marijuana or hashish and then indicate that they had smoked a blunt in the past 30 days. Similarly, respondents could have reported that their last use of marijuana was more than 30 days ago and then indicate use of blunts in the past 30 days. In these situations, respondents were asked why they had reported earlier that they had never used marijuana or hashish or that they had not used it in the past 30 days, respectively.

An important aspect of the processing of variables in this module consisted of assigning codes of 91 and 99 (see Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents said they never smoked part or all of a blunt, they were skipped out of the remaining blunts questions. Thus, respondents were assigned codes of 91 (NEVER USED BLUNTS) to the variables BLNTREC (most recent use of a blunt, corresponding to question BL02) and BLNT30DY (frequency of use of blunts in the past 30 days, corresponding to question BL02A) if they reported that they had never smoked blunts. Similarly, if respondents refused to report whether they had ever smoked part or all of a blunt, that refusal code was assigned to BLNTREC and BLNT30DY as well. In addition, if respondents reported being lifetime users of blunts but reported that they last used blunts more than 30 days ago, question BL02a would have been skipped. Therefore, BLNT30DY was assigned a code of 93 (DID NOT USE BLUNTS IN THE PAST 30 DAYS) in this situation.

Minimal editing was done to the variables BLNT30C1, BLNT30C2, RSNOMRJ, and RSNMRJMO, corresponding to questions BL03, BL04, BL05, and BL06, respectively; these variables pertained to the above-mentioned consistency checks between respondents' answers in the blunts and marijuana modules. If respondents reported not using blunts in the past month or reported that they were a lifetime marijuana user, questions BL03 and BL05 would have been skipped. Therefore, BLNT30C1 and RSNOMRJ were assigned a code of 99 in this situation. Similarly, if respondents were not past month users of blunts or reported last using marijuana in the past 30 days in the core marijuana module (i.e., a response in the marijuana module that was consistent with reported use of blunts in the past 30 days), then questions BL04 and BL06 would have been skipped. Therefore, BLNT30C2 and RSNMRJMO were assigned a code of 99 in these situations.

If respondents were routed to question BL03 (because they reported past month use of blunts but previously reported never using marijuana) and then they refused to answer question BL03, a refusal was propagated from BLNT30C1 to RSNOMRJ. If respondents answered BL03 as "don't know," RSNOMRJ retained a code of 98. Similarly, if respondents were routed to BL04 (because they reported past month use of blunts but previously reported using marijuana but not in the past 30 days) and they refused to answer, a refusal was propagated from BLNT30C2 to RSNMRJMO. Respondents answering "don't know" to question BL04 retained a code of 98 in RSNMRJMO.

In addition, since 2005, respondents who were lifetime users of blunts have been asked how old they were when they first used blunts (BL08). If respondents reported first using blunts within 1 year of their current age, they were asked to report the specific year and month when they first used, with the allowable years ranging from 2010 to 2012. If respondents reported first using blunts at their current age, their last birthday was in the current year, and they were interviewed after their last birthday, the CAI program assumed that the first use of blunts occurred in the current year (i.e., 2012). These respondents were asked only for the month when they first used in the current year. The remaining respondents who first used blunts within 1 year of their current age could be routed to one of two possible questions on the specific year they first used. They then were routed to a question to report on the specific month they first used blunts in the year they had reported previously.

Because the routing logic to the different versions of the year-of-first use and month-of-first-use questions was mutually exclusive, composite sets of year-of-first-use and month-of-first-use variables (BLNTYFU and BLNTMFU, respectively) were created from the individual unedited variables. In addition, the final year-of-first-use variables were recoded to replace codes of 1 or 2 from the questionnaire that denoted specific years with values for the years (i.e., 2010 through 2012). In situations in which the CAI program skipped the year-of-first-use questions because it could be logically inferred that first use of blunts occurred in the interview year (i.e., 2012), a code of "2012" was assigned to BLNTYFU; this was done even if respondents did not know what month they first used in the current year, or if they refused to report what month they first used in the current year. If the year- and month-of-first-use questions had been skipped because respondents first used blunts at ages that were more than 1 year below their current ages, legitimate skip codes were assigned to the BLNTYFU and BLNTMFU variables.

A consistency check also was included for situations in which the apparent age when respondents first used blunts that was calculated from their year and month of first use and their date of birth (CAI variable MYR1STBL) was inconsistent with the age that respondents directly reported for when they first used blunts (CAI variable AGE1STBL). For example, the consistency check was triggered if a 16-year-old respondent reported first using blunts at age 16 but then reported first using blunts in a year and month that would have meant the respondent was 15 years old when he or she first used blunts. No further editing needed to be done if respondents indicated twice in a row that the age at first use that was calculated from the month and year of first use was correct. The CAI program updated the value for AGE1STBL (corresponding to the edited variable BLNTAGE) to agree with the values for the month and year of first use.

A calculation of an age at first use based on respondents' reported year and month of first use of blunts was not done if respondents reported that they first used blunts in the same month as their birth dates; in this situation, a unique age at first use could not be determined. Similarly, this consistency check was not triggered if respondents had missing data in either of the year or month questions, such as if respondents knew the year when they first used blunts but they did not know the month of first use.

If respondents indicated at some point in the consistency check sequence that the value they had originally reported for their age at first use was correct (and by extension, that the year and month of first use was not correct), they had an opportunity to revise the values for their year of first use and their month of first use. These revised reports for year and month of first use were used in subsequent editing steps. Otherwise, respondents' answers to the original year- and month-of-first-use questions were used in subsequent editing.

Table B.15 in Appendix B discusses edits pertaining to the consistency checks for these incidence variables for blunts. For consistency with how the core incidence data have been edited, the default when a respondent did not resolve an inconsistency between the age at first use and the month and year of first use was to favor the age at first use in subsequent editing decisions. Unlike the core incidence data, however, the incidence variables for blunts did not undergo subsequent statistical imputation.

Because these incidence variables were not present in 2004, editing of BLNTREC beyond 2004 did not take into account these incidence data for consistency with how BLNTREC was edited in 2004. Keeping the editing procedures comparable with those in 2004 would permit analysts to examine trends since 2004. However, a flag variable was created that indicated when incidence data suggested more recent use than what the respondent indicated in BLNTREC. The flag was coded as 98 (blank) if no inconsistency existed between BLNTREC and the incidence data.

Table B.16 presents additional edits pertaining to variables in the blunts module. Many of these edits were developed when the module was added to the survey in 2004, particularly with regard to the editing of BLNTREC. For example, if respondents reported using blunts but they also reported never using marijuana, then question BL03 was triggered only when respondents reported using blunts in the past 30 days. Respondents who verified that they had never used marijuana would then have their BLNTEVER answer assigned to 4 (No LOGICALLY

ASSIGNED) and BNLTREC assigned to 81 (NEVER USED BLUNTS, Logically assigned). Table B.16 also discusses miscellaneous edits for the incidence variables that have been present since 2005, other than those described in Table B.15 that were related to consistency checks.

No editing was done to make the incidence data for blunts consistent with incidence data for cigars or marijuana from their respective core sections of the interview, or vice versa. Thus, for example, incidence data in the blunts module could indicate that respondents first smoked a cigar with marijuana in it at an earlier age than they reported for when they first used cigars or marijuana. However, variables in the blunts section of the National Survey on Drug Use and Health (NSDUH) codebooks included a standard footnote to indicate that these noncore data may be inconsistent with data from core modules.

7.4.4 Substance Dependence and Abuse Module

The substance dependence and abuse module asked about symptoms of dependence or abuse in the past 12 months that were associated with the use of alcohol, marijuana, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription psychotherapeutic drugs (i.e., pain relievers, tranquilizers, stimulants, and sedatives). This section also included items to assess for dependence on cigarettes if respondents reported use of cigarettes in the past 30 days. Details on how respondents were classified as having dependence or abuse for alcohol or illicit drugs (i.e., not including cigarettes) are found in Section B.4.2 in Appendix B of the 2012 summary of national findings (Center for Behavioral Health Statistics and Quality, 2013b).

7.4.4.1 Editing of Substance Dependence and Abuse Variables for Nonusers in the Period of Interest

An important aspect of the processing of variables in this module consisted of assigning codes of 91 or 93 (see Section 2.4.2) to variables that had been skipped because respondents never used the substance or they used it but not in the period of interest (i.e., more than 30 days ago for cigarettes or more than 12 months ago for other substances).[106] For alcohol and marijuana, respondents who had used these substances in the past 12 months also were skipped out of the corresponding dependence and abuse questions if they were only infrequent users of these two drugs in the past 12 months. If recency-of-use variables for the psychotherapeutic drugs were assigned a code of 81 because they only drugs they ever used nonmedically were OTC drugs (see Section 6.2.1.5), then any data in the substance dependence and abuse module for that psychotherapeutic drug were overwritten with codes of 81. For cocaine, heroin, and stimulants, however, respondents' answers in the substance dependence and abuse module were retained if they were routed into that respective section in the substance dependence and abuse module because they reported past year use in the special drugs module (see Section 7.4.1.2).

[106] For cocaine, heroin, and stimulants, respondents were not asked the corresponding questions in the substance dependence and abuse module if there was no indication of use in the past 12 months either in the relevant core module (or modules, in the case of cocaine and crack) or in respondents' answers from the special drugs module. However, respondents who did not indicate past year use of cocaine, heroin, or stimulants in the relevant core sections but indicated past year use in the special drugs module *were* routed by the CAI instrument into the relevant drug dependence or abuse questions. For stimulants, this routing logic included situations in which respondents reported use of methamphetamine in the past year in the methamphetamine follow-up items that were added to the special drugs module in 2005 and the follow-up questions that were added in 2006 (see Section 6.2.6).

Special situations where data in the edited variables for stimulants or sedatives were overwritten with codes of 83 are discussed in Section 7.4.4.2.

For alcohol and marijuana, the final, edited 12-month frequency variables (ALCYRTOT and MJYRTOT, respectively) also were used in assigning codes of 93 or 83 to the substance dependence and abuse variables pertaining to these substances. For example, if the edited variable ALCYRTOT indicated that respondents had used alcohol in the past 12 months but on fewer than 6 days in that period, the edited substance dependence and abuse variables for alcohol were assigned codes of 93 if they had been skipped. If respondents answered one or more dependence or abuse questions for alcohol but the final value for ALCYRTOT indicated that they had used alcohol on fewer than 6 days in the past 12 months, the previous answers in the dependence and abuse questions were overwritten with codes of 83. Similar edits were done for marijuana if MJYRTOT indicated that respondents used marijuana on fewer than 6 days in the past 12 months.

7.4.4.2 Editing of Substance Dependence and Abuse Variables for Alcohol through Sedatives

A second important aspect of the processing of the substance dependence and abuse variables for alcohol through sedatives involved assignment of legitimate skip codes of 99 (see Section 2.4.2) when respondents qualified for being asked dependence or abuse questions about a given substance but they legitimately skipped out of one more questions about that substance. For example, the symptom of tolerance to the effects of alcohol was measured through two related questions, DRALC06 ("During the past 12 months, did you need to drink more alcohol than you used to in order to get the effect you wanted?") and DRALC07 ("During the past 12 months, did you notice that drinking the same amount of alcohol had less effect on you than it used to?"). An affirmative answer to either question would indicate tolerance. Thus, if respondents had already answered DRALC06 as "yes," there was no need to ask DRALC07. If the edited variable corresponding to question DRALC06 (ALCNDMOR) was coded as 1 (i.e., "yes"), the edited variable corresponding to question DRALC07 (ALCLSEFX) was assigned a legitimate skip code.

Aside from assignment of codes of 91, 93, or 99, minimal additional editing was done to the substance dependence and abuse variables. However, one notable change was that since 2005, the CAI logic for questions related to stimulant dependence or abuse have taken into account the new follow-up questions SD17A or SD18A that were added to the special drugs module for respondents who had not previously reported methamphetamine use in the core stimulants module (see Section 6.2.6). Thus, if respondents in 2005 indicated past year use of methamphetamine in these noncore questions, they were asked questions in the substance dependence and abuse module about stimulant dependence or abuse. In addition, since 2006, this logic has skipped respondents out of the dependence and abuse questions for stimulants if they reported methamphetamine use in the past year in the special drugs module but then they reported in question SD17A1 that their earlier report of lifetime nonuse of methamphetamine from the core stimulants module was actually the correct response.

Since 2008, the instrument routing logic also has taken into account reports from the special drugs module of past year nonmedical use of Adderall® for asking the stimulant

dependence and abuse questions and past year nonmedical use of Ambien® for asking the dependence and abuse questions for sedatives. In addition, respondents who reported past year use of ketamine, DMT, AMT, "Foxy," or *Salvia divinorum* have been asked the hallucinogen dependence and abuse questions.

In editing the stimulant dependence and abuse variables, stimulant dependence or abuse data were retained for these respondents who indicated past year methamphetamine use or past year nonmedical use of Adderall® in the noncore special drugs items. Consequently, if the stimulant dependence and abuse variables had been skipped because lifetime nonuse of methamphetamine had been logically inferred for the noncore methamphetamine recency MTHAREC and ADDEREC also indicated lifetime nonuse, the edited stimulant dependence and abuse variables were assigned codes of 91 (NEVER USED STIMULANTS). If MTHAREC or ADDEREC indicated past year use, data were retained in the stimulant dependence and abuse variables, even if the other stimulant variables that existed prior to 2005 (including the methamphetamine and other stimulant needle use variables that were in the special drugs module prior to 2005) would have indicated that the respondent was not a past year stimulant user. Similar principles since 2008 applied to the editing of the sedative dependence and abuse variables based on AMBIREC and the hallucinogen dependence and abuse variables based on KETAREC, TRYPREC, and SALVREC (see Section 7.4.1.4). In the case of hallucinogens, for example, data were retained in the dependence and abuse variables if there was any indication of past year use in KETAREC, TRYPREC, or SALVREC, even if the core variable HALLREC did not indicate past year use.

As noted previously, substance dependence and abuse data for psychotherapeutic drugs were overwritten if the only drugs that respondents used nonmedically were OTC drugs. For stimulants and sedatives, the data were overwritten with codes of 83 rather than with codes of 81 if ADDEREC or AMBIREC, respectively, indicated nonmedical use more than 12 months ago.

Therefore, in addition to the flag variable STMDAFLG that has been present since 2005 to indicate differences for analysts in the editing of the stimulant variables in the substance dependence and abuse module, the flag variables HALDAFLG and SEDDAFLG were created to indicate differences in how the hallucinogen and sedatives variables in the substance dependence and abuse module have been edited since 2008. In the case of STMDAFLG, levels 1 through 6 have been present since 2005 to document the effects of the noncore methamphetamine data on the editing of the stimulant dependence and abuse variables, and levels 11 through 16 have been present since 2008 to document the effects of ADDEREC on the editing of these variables. Documentation for these levels in STMDAFLG in 2012 was as follows, where "Now" refers to this survey year (i.e., 2012):

1 = Never used based on 04/Now: Not past yr user

2 = Not past yr user based on 04/Now: Blank

3 = Never used based on 04/Now: Blank

4 = Blank based on 04/Now: Past yr user

5 = Not past yr user based on 04/Now: Past yr user

6 = Never used based on 04/Now: Past yr user

11 = Never used based on 07/Now: Not past yr user

12 = Not past yr user based on 07/Now: Blank

13 = Never used based on 07/Now: Blank

14 = Blank based on 07/Now: Past yr user

15 = Not past yr user based on 07/Now: Past yr user

16 = Never used based on 07/Now: Past yr user

98 = BLANK (NO DIFFERENCE BETWEEN 04 AND NOW)

Thus, a code of 98 in STMDAFLG meant that the presence of the noncore methamphetamine and Adderall® items in the special drugs module had no effect on how the stimulant dependence and abuse variables have been edited since 2004. For example, if respondents indicated past year nonmedical use of stimulants in the core stimulants module (i.e., but they did not indicate methamphetamine use) or if they indicated that they used stimulants with a needle in the past year (from MTNLDREC or OSTNLREC in the special drugs module), they would have been asked questions about stimulant dependence or abuse regardless of how they answered the methamphetamine and Adderall® questions in the special drugs module. Similarly, if respondents were classified as lifetime nonusers of stimulants based on their answers in the core stimulants module, and the special drugs data (including the additional methamphetamine Adderall® items) continued to indicate that these respondents never used methamphetamine or other prescription-type stimulants (or Adderall®), the stimulant dependence and abuse items were coded as 91, as would have been the case in 2004 and earlier years.

For codes 1 through 6 in STMDAFLG since 2008, ADDEREC did not indicate past year nonmedical use of Adderall®. A code of 1 in STMDAFLG meant that respondents who skipped the stimulant dependence and abuse questions also would have skipped out of these items in 2004 and earlier years. The only difference was that in 2004 (and earlier), the edited stimulant dependence and abuse variables would have been coded as 91 (NEVER USED STIMULANTS). Based on data from the additional methamphetamine variables in the special drugs module, these variables currently were coded as 93 (DID NOT USE STIMULANTS IN THE PAST 12 MONTHS).

A code of 2 meant that respondents would have been coded as lifetime but not past year users of stimulants based solely on the variables that existed in 2004. The stimulant dependence and abuse variables currently were coded as 98 (blank) because these respondents reported methamphetamine use in the additional special drugs items, but it was not clear whether they used methamphetamine in the past year.

A code of 3 meant that respondents would have been coded as nonusers based solely on the variables that existed in 2004. The stimulant dependence and abuse variables were coded as 98 (blank) for the same reason as that given for situations in which STMDAFLG was coded as 2.

Beginning with codes of 4, nonmissing data existed in the current stimulant dependence and abuse variables that would not have existed in 2004.

- A code of 4 meant that the variables that existed in 2004 indicated that respondents were at least lifetime nonmedical users of stimulants, but it was not clear whether they had used in the past year. However, these respondents indicated past year methamphetamine use in the additional special drugs items.
- A code of 5 meant that respondents would have been classified as lifetime but not past year nonmedical users of stimulants based on the variables that existed in 2004, but they indicated past year methamphetamine use in the additional special drugs items.
- A code of 6 indicated the greatest potential difference between 2004 and currently. Based on the variables that existed in 2004, these respondents would have been classified as never having used stimulants nonmedically, but they indicated past year methamphetamine use in the additional special drugs items.

Levels 11 through 16 in STMDAFLG were analogous to levels 1 through 6. For levels 11 through 16, however, the noncore methamphetamine data would not have affected how the stimulant dependence and abuse variables were edited but ADDEREC did. For example, a code of 11 in STMDAFLG meant that respondents who skipped the stimulant dependence and abuse questions in 2008 onward also would have skipped out of these items in 2007, even with the noncore methamphetamine data. The difference was that in 2007, the edited stimulant dependence and abuse variables would have been coded as 91 (NEVER USED STIMULANTS) because the respondent was classified as a lifetime nonuser of stimulants based on the core stimulant data and noncore methamphetamine data. However, ADDEREC indicated that these respondents used Adderall® nonmedically more than 12 months ago but in their lifetime. Consequently, the stimulant dependence and abuse variables in this situation were assigned a code of 93 (DID NOT USE STIMULANTS IN THE PAST 12 MONTHS); this was analogous to the situation where STMDAFLG = 1.

Similarly, since 2008, codes 14 through 16 indicated that nonmissing data existed in the stimulant dependence and abuse variables that would not have existed in 2007. In particular, a code of 16 indicated the greatest potential difference between 2007 and currently. Based on the variables that existed in 2007, these respondents would have been classified as never having used stimulants nonmedically (including never having used methamphetamine based on the noncore special drugs data). However, they indicated past year nonmedical use of Adderall®.

In HALDAFLG and SEDDAFLG, levels 1 through 6 corresponded to levels 1 through 6 in STMDAFLG. In HALDAFLG and SEDDAFLG, however, data were compared relative to the results of how the relevant dependence and abuse variables would have been edited in 2007 (rather than 2004). For example, a code of 1 in HALDAFLG meant that respondents who skipped the hallucinogen dependence and abuse questions in 2008 onward also would have skipped out of these items in 2007. The difference was that in 2007, the edited hallucinogen dependence and abuse variables would have been coded as 91 (NEVER USED HALLUCINOGENS) because the core hallucinogen data indicated that the respondents never used hallucinogens. However, there was some indication of use more than 12 months ago in KETAREC, TRYPREC, or SALVREC but no indication of use in the past 12 months. Consequently, the current hallucinogen dependence and abuse variables were assigned a code of 93 (DID NOT USE HALLUCINOGENS IN THE PAST 12 MONTHS) in this situation.

7.4.4.3 Editing of Cigarette Dependence Variables

The questionnaire items that were used to determine cigarette dependence (also known as "nicotine dependence") in the 2012 NSDUH have been included in the survey since 2001. The method for determining dependence involved calculating a continuous score from the Nicotine Dependence Syndrome Scale (NDSS) (Shiffman, Hickcox, Gnys, Paty, & Kassel, 1995; Shiffman, Waters, & Hickcox, 2003). The score was calculated from the 17 questions shown in Table 7.1 that were asked of respondents who used cigarettes in the past 30 days. For each of these items, respondents who had used cigarettes in the past 30 days were asked to choose an answer from the following scale: 1 = Not at all true; 2 = Somewhat true; 3 = Moderately true; 4 = Very true; or 5 = Extremely true. For details on how the estimate for cigarette dependence was calculated based on the NDSS, see Section B.4.2 in Appendix B of the 2009 summary of national findings (Office of Applied Studies, 2010a; 2010b).

As for other variables in the substance dependence and abuse module, editing of the cigarette dependence variables also involved assignment of legitimate skip codes of 99 according to the skip logic within this section of the module. For example, question DRCGE06A (not shown in Table 7.1) asked if respondents who reported cigarette use in the past 30 days had any friends who did *not* smoke cigarettes. If DRCGE06A was answered as "no," question DRCGE06B was skipped (i.e., choosing at times not to be around friends who do not smoke because they do not like it when the respondent smokes). The edited variable CIGFNLKE corresponding to DRCGE06B was assigned a code of 99 if DRCGE06A was answered as "no."[107]

Table 7.1 Mapping of Cigarette Dependence Questions to Edited Variables

Question	Edited Variable	Question Text
DRCGE01	CIGIRTBL	After not smoking for a while, you need to smoke in order to feel less restless and irritable.
DRCGE02	CIGCRAVE	When you don't smoke for a few hours, you start to crave cigarettes.
DRCGE03	CIGCRAGP	You sometimes have strong cravings for a cigarette where it feels like you're in the grip of a force you can't control.
DRCGE04	CIGINCTL	You feel a sense of control over your smoking – that is, you can "take it or leave it" at any time.
DRCGE05	CIGAVOID	You tend to avoid places that don't allow smoking, even if you would otherwise enjoy them.
DRCGE07	CIGPLANE	Even if you're traveling a long distance, you'd rather not travel by airplane because you wouldn't be allowed to smoke.
DRCGE08	CIGRNOUT	You sometimes worry that you will run out of cigarettes.
DRCGE09	CIGREGDY	You smoke cigarettes fairly regularly throughout the day.
DRCGE10	CIGREGWK	You smoke about the same amount on weekends as on weekdays.

[107] Consistent with the principles in Section 2.4.2, CIGFNLKE retained missing values if DRCGE06B was skipped because respondents answered DRCGE06a as "don't know" or "refused."

Table 7.1 Mapping of Cigarette Dependence Questions to Edited Variables (continued)

Question	Edited Variable	Question Text
DRCGE11	CIGREGNM	You smoke just about the same number of cigarettes from day to day.
DRCGE12	CIGNMCHG	It's hard to say how many cigarettes you smoke per day because the number often changes.
DRCGE13	CIGSVLHR	It's normal for you to smoke several cigarettes in an hour, then not have another one until hours later.
DRCGE14	CIGINFLU	The number of cigarettes you smoke per day is often influenced by other things – how you're feeling or what you're doing, for example.
DRCGE15	CIGNOINF	Your smoking is not affected much by other things. For example, you smoke about the same amount whether you're relaxing or working, happy or sad, alone or with others.
DRCGE16	CIGINCRS	Since you started smoking, the amount you smoke has increased.
DRCGE17	CIGSATIS	Compared to when you first started smoking, you need to smoke a lot more now in order to be satisfied.
DRCGE18	CIGLOTMR	Compared to when you first started smoking, you can smoke much, much more now before you start to feel anything.

No additional editing was done to the cigarette dependence variables if respondents reported cigarette use in the past 30 days or they used cigarettes in the past 30 days but skipped out of questions that did not apply. In particular, no editing was done when respondents entered the same response for all items (e.g., keying a "1" to every item). If respondents entered the same response to all cigarette dependence items, however, that would strongly suggest that they were not paying careful attention to the questions. For example, if a respondent chose "1" (i.e., not at all true) in question DRCGE03 (see Table 7.1), the respondent would be expected to choose a response other than "1" to indicate some degree of truth to the statement in question DRCGE04 about having some sense of control over smoking. Nevertheless, when respondents entered the same response to all cigarette dependence items, these data were retained in order to allow analysts to decide how they wanted to handle these cases.

7.4.4.4 Imputation of Cigarette Dependence Variables

Unlike the majority of variables that are imputed for NSDUH, imputation of the cigarette dependence variables was not performed using the PMN methodology. For respondents who used cigarettes in the past 30 days and provided complete data for all 17 of the dependence questions that were used to calculate the NDSS scale value, imputation-revised cigarette dependence variables were simply assigned the values from the corresponding edited variables.

For respondents who used cigarettes in the past 30 days and gave a valid response to 16 of the 17 NDSS items, the predicted mean for the one missing item was obtained using the coefficients corresponding to the other 16 nonmissing covariates from the appropriate weighted least squares regression. For example, if CIGIRTBL was the variable whose missing value was to be imputed, CIGIRTBL would be specified as the dependent variable in the model, and the

remaining 16 NDSS variables served as the covariates: CIGCRAVE, CIGCRAGP, CIGINCTL, CIGAVOID, CIGPLANE, CIGRNOUT, CIGREGDY, CIGREGWK, CIGREGNM, CIGNMCHG, CIGSVLHR, CIGINFLU, CIGNOINF, CIGINCRS, CIGSATIS, and CIGLOTMR. The imputation-revised variable was then set to the predicted mean.[108]

Respondents who used cigarettes in the past 30 days but answered 15 or fewer of the cigarette dependence questions were left out of the modeling process entirely. The missing values for these respondents remained missing in the imputation-revised variables that were derived from the edited NDSS variables.

Across all respondents (regardless of how many of the NDSS questions they answered), no response propensity adjustments were performed for the item respondent weights used in the regression-based imputation models. The analysis weights described in Section 3.3.1 were applied instead.

The NDSS mean value was calculated from imputation-revised versions of the 17 nicotine dependence questionnaire variables. The NDSS mean value was set to missing for respondents who used cigarettes in the past 30 days but answered 15 or fewer of the cigarette dependence questions.

7.4.4.5 Summary Information for Cigarette Dependence Variables

In the 2012 NSDUH, 151 cases had valid values for 16 of the 17 questions. These cases were imputed. In 2012, 136 cases had fewer than 16 valid values for the NDSS scale. For these cases, no imputations were performed. Note that it was possible that the respondent was eligible to answer the questions about cigarette dependence because he or she was imputed to be a past month cigarette user. This situation occurred five times in the 2012 NSDUH.

Table 7.2 summarizes the eligibility of respondents to answer the cigarette dependence questions and reasons why respondents were classified as eligible or not eligible to be imputed. Furthermore, this table provides details about the amount of cigarette dependence data that were missing for eligible respondents. It also provides information on whether the respondent was imputed to be a past month cigarette user. Consequently, the respondent would be eligible to have cigarette dependence data but would have missing data for all the cigarette dependence variables.

Of the 287 eligible respondents in 2012 who did not answer all 17 NDSS questions, the majority (151) were missing a response for only one of the items (Table 7.2). Any respondent with more than one of the 17 items missing (47.4 percent of the eligible cases with incomplete data) did not have his or her missing responses replaced with imputed values, and no cigarette dependence score was calculated for those respondents. For the respondents missing only one response, imputation was used to fill in the values for the missing variable, using information from the other 16 nonmissing variables, through weighted least squares regression models. This resulted in 17 regression-based imputation models, where the response variable for each model

[108] Because the response variable and covariates were treated as continuous in the models, it is possible for a predicted mean, and therefore an imputation-revised value, to exceed five or be less than one.

was the edited variable that corresponded to each item in the NDSS, and the covariates in each model were the remaining NDSS variables.

Table 7.2 Summary of Response Patterns for 2012 NDSS Variables

Number of Valid NDSS Variables	Past Month Smoker	Past Month Smoker Status Imputed	Eligible to Answer NDSS Questions	NDSS Variables Imputed	Frequency	Percentage of Total
N/A	No	No	No	N/A[1]	53,238	77.94
N/A	No	Yes	No	N/A[1]	9	0.01
Subtotal					53,247	77.95
0	Yes	No	Yes	No	14	0.02
0	Yes	Yes	Yes	No	5	0.01
1-15	Yes	No	Yes	No	117	0.17
Subtotal					136	0.20
16	Yes	No	Yes	Yes	151	0.22
17	Yes	No	Yes	N/A[2]	14,775	21.63
Total					68,309	100.00

N/A = not applicable; NDSS = Nicotine Dependence Syndrome Scale.

[1] The respondent was not eligible to answer the NDSS questions.

[2] None of the NDSS variables were missing for this scenario.

7.4.5 Special Topics Module

The special topics module asked about arrests in the respondents' lifetime and in the past 12 months, including arrests for specific offenses in the past 12 months (not counting minor traffic violations). This section also included questions about respondents being on probation or parole in the past 12 months, operating vehicles under the influence of alcohol or illegal drugs in the past 12 months, and respondents' knowledge about their States' marijuana laws.

If respondents reported that they had never been arrested in their lifetime and they did not report being on probation or parole in the past 12 months (see below), the edited variables pertaining to arrests in the past 12 months were assigned legitimate skip codes. Other standard edits described in Section 2.4.3 pertaining to situations where respondents answered "don't know" or "refused" to the lifetime arrest question were applied to the past year arrest variables that had been skipped.

Similarly, if respondents reported being arrested in their lifetime but reported being arrested 0 times in the past 12 months, the questions pertaining to arrests for specific offenses in the past 12 months were assigned legitimate skip codes. Respondents who did not know how many times they were arrested in the past 12 months or who refused to answer this question were asked whether they were arrested for specific offenses in the past 12 months. This was consistent with the logic in 1999 and since 2001 but differed from the logic in 2000, when respondents who answered "don't know" or "refused" to the question about the number of specific arrests in the

past 12 months were skipped out of questions about arrests for specific offenses in the past 12 months.

Respondents also were skipped out of questions pertaining to driving under the influence of alcohol or illegal drugs if they reported in the core modules that they never used alcohol, marijuana, cocaine, heroin, hallucinogens, inhalants, or prescription psychotherapeutics for nonmedical reasons; this included situations in which respondents were classified as having never used stimulants nonmedically based on their answers in the core stimulants module, and the special drugs data (including the added special drugs items on use of methamphetamine) continued to indicate that the respondents had never used stimulants. When respondents were lifetime nonusers of alcohol and the other drugs mentioned above, all skipped variables pertaining to driving under the influence of alcohol or illegal drugs were assigned a code of 91 to indicate that the respondents were lifetime nonusers of all of these substances (Section 2.4.2). If respondents were skipped out of one or more of the substance use and driving items because their most recent use of a drug was more than 12 months ago, the edited variables were assigned legitimate skip codes.

In addition, respondents were asked about driving under the influence of illegal drugs (alone or in combination with alcohol) if they reported use of methamphetamine in the past year in the methamphetamine follow-up items from the special drugs module. No editing was done to the variables about driving under the influence of illegal drugs in this situation. However, the additional methamphetamine items from the special drugs module would have affected routing to the items in special topics about driving under the influence of illegal drugs only if these methamphetamine items were the sole indication of use of drugs other than alcohol in the past year. Stated another way, if respondents had reported past year use of marijuana, cocaine, heroin, hallucinogens, or inhalants, or if they had reported past year nonmedical use of pain relievers, tranquilizers, stimulants other than methamphetamine, or sedatives, they would have been asked the questions about driving under the influence of illegal drugs regardless of how they answered the questions from the special drugs module about methamphetamine. A flag (DRVFLAG) was created for use by the analysis team to identify any respondents whose only use of drugs other than alcohol came from the methamphetamine items in the special drugs module; updated logic for this flag has been in place since 2006 to take into account the consistency checks between the core and noncore responses that were added to the special drugs module in 2006 (see Section 6.2.6.

The variable pertaining to a State's maximum legal penalty for first offense possession of an ounce or less of marijuana was renamed from MXMJPENL to MXMJPNLT. The variable name was changed to reflect the addition of a response category for "No penalty" and to alert analysts that data in MXMJPNLT from 2006 onward may not be comparable with data in MXMJPENL prior to 2006.

Minimal editing was done to the variables MXMJPNLT, MXMJSURE, and MEDMJUSE, corresponding to questions SP07, SP08, and SP09, respectively; these variables pertained to knowledge about State marijuana laws and penalties. If respondents indicated in MXMJPNLT that they did not know the maximum legal penalty in their States for possession of an ounce or less of marijuana for personal use, or if they refused to answer this question, the edited variable MXMJSURE (regarding respondents' degree of certainty about their answer to

question SP07) was assigned a legitimate skip code. In addition, if interviewers had entered incorrect information in the FIPE4 question regarding the State where the respondent's sampled dwelling unit was located, the variables MXMJPNLT, MXMJSURE, and MEDMJUSE were assigned bad data codes.[109] This latter edit was done because the State that respondents were asked about in these questions was governed by the State that interviewers entered in FIPE4. Hence, if interviewers entered incorrect State information in FIPE4, the answers that respondents provided in questions SP07 through SP09 were deemed to be questionable. For example, if a respondent lived in California (FIPE4 = 5) but the interviewer entered that the respondent's sampled dwelling unit was in Colorado (FIPE4 = 6), the respondent would be asked for information on marijuana laws in Colorado.

Table B.17 in Appendix B discusses additional edits that were specific to the special topics module. For example, respondents could report that they had never been arrested in their lifetime but could report that they were on probation, parole, or supervised release in the past 12 months. Because someone could not be on probation or parole without first having been arrested for a crime, these respondents were logically inferred to have been arrested in their lifetime. When this situation occurred, the skipped variables pertaining to arrests in the past 12 months retained a value of blank.

Since 2002, respondents who reported in question SP02 that they were arrested and booked for breaking the law at least once in the past 12 months and gave negative answers to every question about specific arrests (including arrests for "some other offense") have been routed to a follow-up question, SP03R, that asked respondents to verify their previous answer from SP02. If respondents did not indicate that their previous answer from question SP02 was correct, they were routed to question SP03S, where they were allowed to update their answer regarding the number of times they were arrested or booked in the past 12 months. If respondents indicated in SP03S that they had been arrested or booked 0 times in the past 12 months, the edited variable NOBOOKYR was assigned a value of 0. Further, when SP03s indicated that respondents had been arrested and booked 0 times in the past 12 months, it was logically inferred that all items pertaining to specific arrests in the past 12 months should have been skipped. Therefore, all of the variables associated with specific arrests in the past 12 months were assigned codes of 89.

If respondents indicated in SP03R that their previous answer from SP02 was correct, or if they reported being arrested and booked for at least one offense in SP03s, they were asked to specify at least one offense for which they were arrested and booked in the past 12 months (questions SP03RSP or SP03SSP). If respondents specified a legitimate offense in SP03RSP or SP03SSP after giving negative answers to every question about specific arrests, the "OTHER, Specify" variable BKOTHOFF was updated to incorporate the response from SP03RSP or SP03SSP. A code of 3 (Yes LOGICALLY ASSIGNED) also was assigned to the "some other offense" variable (BKOTH) to indicate that this edit had taken place (see Section 2.4.5).

Any information that respondents specified in SP03RSP or SP03SSP also was used to edit variables pertaining to offenses that respondents previously had been asked about. For example, if respondents had answered question SP03A as "no" (i.e., had not been arrested and

[109] Creation of the edited variable STATELOC from FIPE4 is discussed in detail in Chapter 4.

booked for motor vehicle theft in the past 12 months) but then specified in SP03RSP or SP03SSP that motor vehicle theft was one of the offenses for which they had been arrested, the edited variable BKMVTHFT was assigned a code of 3.

If respondents reiterated in SP03R or SP03S that they had been arrested and booked for at least one offense in the past 12 months but still did not report a legitimate offense in questions SP03RSP or SP03SSP (including situations in which they answered SP03RSP or SP03SSP as "don't know" or "refused"), then a code of 5 was assigned to BKOTH. This code of 5 had the following meaning: 5 = Offense unknown. Stated another way, the response from SP02 or SP03S was retained in NOBOOKYR to indicate that the respondents were arrested in the past 12 months, but it was not possible to determine the specific crime for which they were arrested.

7.4.6 Marijuana Purchases Module

The marijuana purchases module focused on the acquisition of marijuana. Administration of questions in this module was limited to respondents who had reported previously that they used marijuana in the past 12 months. These respondents were asked how they obtained the last marijuana they used, including buying it, trading something for it, getting it for free (or sharing someone else's), or growing it. The module also included questions about the contexts in which respondents engaged in transactions involving marijuana, including where respondents were when they bought, traded for, or got marijuana for free; from whom the respondents got the marijuana (if they did not grow it themselves); and whether they sold or gave away any of this marijuana (including those respondents who grew their own).

If respondents did not report buying the last marijuana they used, they were asked a follow-up question to identify those who had bought any marijuana in the past 12 months. Respondents who reported purchasing the last marijuana they used or who reported purchasing it at any time during the past 12 months were asked more detailed questions about their purchases of marijuana, and they were skipped out of questions pertaining to trading for marijuana, getting it for free, or growing it.

Similarly, respondents who reported that they traded something for the last marijuana they used and who had not bought marijuana at any time during the past 12 months were asked more detailed questions about trading for marijuana. If respondents did not report trading for the last marijuana they used, they were asked a follow-up question to identify those who had traded something for marijuana in the past 12 months. Respondents who had not been routed into questions about buying marijuana and who were asked more detailed questions about trading for marijuana were skipped out of questions pertaining to getting marijuana for free or growing it.

Respondents who were routed into more detailed questions about purchases of marijuana were asked whether they last bought marijuana in "joints" or in loose form, the quantity they purchased the last time they bought marijuana, and the price they paid. Similar questions were asked of respondents who were routed into questions about trading for marijuana, except that these respondents were asked to estimate the worth of the marijuana they obtained through trading.

Edits in this module principally involved assigning appropriate legitimate skip codes based on the logic for determining whether respondents should be administered the module, or the routing logic within the module, if respondents had used marijuana in the past 12 months (Section 2.4.2). If respondents reported in the marijuana module in the core that they had never used marijuana, the edited variables in the marijuana purchases module were assigned codes of 91 (or 991, etc.) to indicate that respondents had skipped out of the module because they were lifetime nonusers of marijuana. Similarly, if respondents' edited marijuana recency MJREC indicated that they last used marijuana more than 12 months ago, the edited variables in the marijuana purchases module were assigned codes of 93 (or 993, etc.) to indicate that respondents had skipped out of the module because they had used marijuana, but not in the past year. If respondents had been skipped out of the marijuana purchases module but their edited marijuana recency had a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED), the skipped marijuana purchases variables retained codes of blank because at least some of these respondents potentially used marijuana in the past 12 months and would have been eligible to be asked questions in the marijuana purchases module.

If respondents previously reported that they had used marijuana in the past 12 months, a key aspect of the editing of variables in the marijuana purchases module involved assignment of legitimate skip codes (99, 999, etc.) according to how respondents were routed through the module. As discussed previously, for example, respondents who gave some report of having bought marijuana were skipped out of questions about trading for marijuana, growing it, or getting it for free. Similarly, respondents who gave some indication of having traded for marijuana (without having indicated buying it) were skipped out of questions related to growing it or getting it for free. If respondents reported buying or trading for marijuana and bought or traded for it in joints, they were skipped out of questions pertaining to buying or trading for marijuana in loose form, and vice versa. If respondents bought or traded for marijuana in loose form, respondents also were routed into or skipped out of questions about the quantities they obtained based on whether they reported purchasing or trading for grams, ounces, or pounds of marijuana. In addition, respondents who reported that they grew the last marijuana they used (without having indicated that they bought or traded for marijuana) were skipped out of questions related to getting marijuana for free, and respondents who reported that they got their last marijuana for free were skipped out of questions related to growing it.

The remaining processing of the variables in the marijuana purchases module involved creating summary variables for the price that respondents paid for the last marijuana they bought or the estimated value of the marijuana they got through a trade. Respondents were first asked to report broad categories of prices. For some of these broader categories (e.g., if respondents reported paying $21.00 to $50.99), respondents were asked to report more detailed price categories (e.g., $21.00 to $30.99; $31.00 to $40.99; $41.00 to $50.99) in order to define more narrowly how much they paid for the marijuana (or how much they estimated the marijuana to be worth). The routing to these more detailed questions was contingent on the broader price category that respondents reported, such that responses to the more detailed price questions were mutually exclusive. Therefore, "composite" summary cost variables were created based on this routing logic.

For example, if respondents reported buying marijuana in loose form the last time, the broad price category variable was called MMLSPCTB (corresponding to question MJE20),

where "LS" stood for marijuana in loose form, and "PCTB" stood for "broad price category." Similarly, the detailed price category variable for buying marijuana in loose form was called MMLSPCAT and was derived from responses in questions MJE20 through MJE25. If, for example, a respondent reported in question MJE20 that he or she paid "$21.00 to $50.99" for the last marijuana purchase (level 4 in question MJE20), MMLSPCAT was coded as 41 if the respondent reported paying $21.00 to $30.99 (level 1 in question MJE21); 42 if the respondent reported paying $31.00 to $40.99 (level 2 in question MJE21); and 43 if the respondent reported paying $41.00 to $50.99 (level 3 in question MJE21).

If respondents reported a broad price category for the marijuana they bought or traded for but they did not know (or refused to report) a more detailed price, the response from the "broad" price category variable (e.g., MMLSPCTB) was used to create a value for the corresponding detailed price category variable (e.g., MMLSPCAT). For example if respondents reported paying $21.00 to $50.99 in question MJE20 but they did not recall more detailed information, the variable MMLSPCAT was assigned a code of 40. This code indicated that it could at least be determined that the respondent paid $21.00 to $50.99, but that more detailed information was not available.

In addition, respondents who reported that they bought marijuana, traded something for marijuana, or got it for free in the past 12 months were asked where they were the last time they got marijuana in these different ways (edited variables MMBPLACE, MMTPLACE, and MMFLACE for where respondents were when they last bought, traded for, or got marijuana for free, respectively). Available response categories in these items were (1) inside a public building, such as a store, restaurant, sports arena, bar, or club; (2) inside a school building; (3) outside on school property; (4) inside a home, apartment, or dorm; (5) outside in a public area, such as a parking lot, street, or park; or (6) some other place.

"OTHER, Specify" variables have been included in the marijuana purchases module since 2005 to capture further information from respondents who reported that they were in some other place when they last bought marijuana, last traded something for marijuana, or last got marijuana for free. If the other place that respondents specified corresponded to one of the response categories mentioned above, the relevant category in the "place" variable was assigned, plus a value of 10 (e.g., in MMBPLACE for buying marijuana, corresponding to question MJE27). Suppose, for example, that respondents reported that they were in some other place when they last bought marijuana (i.e., category 6 chosen in question MJE27), but they specified something that corresponded to them being outside in a public area. That would have been equivalent to respondents having selected category 5 from MJE27. Therefore, MMBPLACE was assigned a code of 15, where 15 = Outside in a public area LOGICALLY ASSIGNED.

7.4.7 Prior Substance Use Module

The prior substance use module covered a wide variety of topics:

- retrospective use of marijuana, cigarettes, alcohol, and cocaine in the year prior to the past 12 months (also referred to below as the past 12 to 24 months);
- the age, year, and month when respondents last used substances covered in the core section of the interview (i.e., age of last use [ALU], year of last use [YLU], and

month of last use [MLU]) or in the noncore special drugs module for methamphetamine if they were lifetime but not past month users of these substances;

- sources of psychotherapeutic drugs that respondents used nonmedically in the past month or past year, including how friends or relatives obtained psychotherapeutic drugs that respondents subsequently got from these other persons; and
- the sequence of initiation of use of cigarettes, alcohol, and marijuana.

One of the important aspects of the processing of variables in this module consisted of assigning codes of 91 and 99 (Section 2.4.2) to variables that had been skipped because the questions did not apply. For example, if respondents never used marijuana or indicated use in the past 30 days, they were skipped out of the questions asking for their age, year, and month when they last used marijuana. Respondents were skipped out of the questions pertaining to the age, year, and month when they last smoked cigarettes daily if (1) they smoked every day in the past 30 days, or (2) they never had a period in their lives when they smoked cigarettes every day for at least 30 days. When a given core recency variable (e.g., MJREC for marijuana) had a refusal code and the corresponding variables in the prior substance use module had been skipped, that refusal from the core recency was propagated to the edited prior substance use variables (Section 2.4.3). Documentation of edits for specific subsections of the prior substance use module is presented in the remainder of this section.

7.4.7.1 Retrospective Use

Since 2003, this module has included a question about retrospective use of marijuana in the year prior to the past 12 months (i.e., edited variable MRJYRBFR, corresponding to question LU01). In addition, retrospective questions on use of cigarettes, alcohol, and cocaine in the year before the past 12 months (edited variables CIGYRBFR, ALCYRBFR, and COCYRBFR, respectively) have been included since 2005. These questions were analogous to MRJYRBFR.

If respondents never used the drug of interest (e.g., marijuana), the retrospective variable (e.g., MRJYRBFR) was assigned a code of 91. In addition, the cigarette recency variable CIGREC indicated if respondents last smoked cigarettes more than 3 years ago (CIGREC = 4). Logically, if respondents last smoked cigarettes more than 3 years ago, they would not have smoked a cigarette in the past 12 to 24 months. Therefore, when CIGREC indicated that respondents last smoked cigarettes more than 3 years ago and question LU37 (corresponding to CIGYRBFR) had been skipped, CIGYRBFR was assigned a code of 99 (LEGITIMATE SKIP).

The retrospective variables MRJYRBFR, ALCYRBR, CIGYRBFR, and COCYRBFR were not edited for consistency with other data for these drugs in the prior substance use module. For example, if MRJYRBFR indicated that respondents used marijuana in the past 12 to 24 months but respondents indicated that they last used marijuana at an age in edited variable MRJAGLST that would suggest that they last used marijuana more than 2 years ago, no editing was done to either MRJYRBFR or MRJAGLST. However, a codebook "NOTE" was added to alert analysts that these types of inconsistencies could exist between the related variables for these drugs.

7.4.7.2 Last Use of Drugs

The prior substance use module included questions about the last use of all drugs that were covered in the core section of the interview. If respondents were lifetime users of specific substances in the core section but had not used these substances in the past 30 days, they were asked in this module for the age, year, and month when they last used these drugs or tobacco (i.e., age of last use [ALU], year of last use [YLU], and month of last use [MLU]). If respondents ever had a period of smoking cigarettes daily but had not smoked every day in the past 30 days, they also were asked for the age, the year, and the month when they last smoked cigarettes on a daily basis.

Specifically, respondents who last used a given drug more than 30 days ago[110] were asked how old they were when they last used that drug. If respondents reported last using the drug within 1 year of their current age, they were asked to report the specific month and year when they last used, with the allowable years ranging from 2010 to 2012. If respondents reported last using the drug at their current age and their birth month was earlier than the interview month (i.e., they reached their current age in the same year that they were interviewed), the CAI program assumed that the last use of the drug occurred in the current year (i.e., 2012). These respondents were asked only for the month that they last used in the current year. The remaining respondents who last used a drug within 1 year of their current age could be routed to one of two possible questions on the specific year they last used. They then were routed to a question to report on the specific month that they last used the drug in the year they had reported previously.

Because the routing logic to the different versions of the month- and year-of-last-use questions was mutually exclusive, composite sets of month-of-last-use and year-of-last-use variables were created from the individual unedited variables. In addition, if respondents indicated a specific year that they last used a drug, the final year-of-last-use variables for 2012 were recoded to replace codes from the questionnaire with values for the years (i.e., 2010 through 2012). If respondents confirmed that they last used a drug at their current age and were interviewed subsequent to their birthday, a code of "2012" was assigned to the year of last use; this was done even if respondents did not know what month they last used in the current year, or if they refused to report what month they last used in the current year. If the month- and year-of-last-use questions had been skipped because respondents last used the drug more than 1 year younger than their current ages, legitimate skip codes were assigned to the final month- and year-of-last-use variables.

Since 2005, consistency checks have been included in the module that were triggered when the values for the month and year of last use were inconsistent with the age at last use. Specifically, for respondents who recently stopped use of a given drug, the CAI program calculated a second age at last use based on the month- and year-of-last-use data by comparing these data with the respondent's date of birth. This comparison was not done if the respondent reported last use of the drug in the same month that he or she was born; a unique age at last use could not be determined from the month and year of last use in these situations because it was not known whether the drug use occurred before or after the respondent's birthday. Similarly, a

[110] Subsequent discussion also applies to respondents whose last period of smoking cigarettes every day occurred more than 30 days ago.

consistency check was not triggered if the respondent had missing data in either of the month or year questions, such as if the respondent knew the year when he or she last used a drug but did not know the month of last use.

In the remaining situations in which respondents provided complete data for the month and year of last use, a consistency check was triggered if the month and year of last use suggested that respondents stopped use of the drug at an earlier or a later age than what they had previously reported in their age-at-last-use question. For example, a consistency check was triggered if a 16-year-old respondent reported last using a drug at age 16 but then reported last using the drug in a month and year that would have meant the respondent was 15 years old when he or she last used the drug. No editing needed to be done if respondents indicated twice in a row that the age at last use that was calculated from the month and year of last use was correct. The CAI program updated the value for the age at last use (e.g., AGELSTCG for cigarettes) to agree with the values for the month and year of last use.

If respondents indicated at some point in the consistency check sequence that the value they had reported for their age at last use (e.g., question LU03 for cigarettes) was correct, they had an opportunity to revise the values for their year of last use and their month of last use. If a consistency check was triggered between the age at last use and data in the month and year of last use, the month and year of last use were updated with any year and month data that the respondent entered in the consistency checks (e.g., LUCG07 and LUCG07a for any cigarette use). These data were used in subsequent editing steps. Otherwise, the month- and year-of-last-use data were picked up from the original source variables (e.g., LU03a through LU03d for any cigarette use) for use in subsequent editing.

Table B.18 in Appendix B presents the edits that were implemented when consistency checks were triggered between the age at last use, year of last use, and month of last use. The default when a respondent did not resolve an inconsistency between the age at last use and the month and year of last use was to favor the age at last use in subsequent editing decisions.

Respondents also were skipped out of the year- and month-of-last-use questions if they indicated that they last used a drug (or last smoked cigarettes every day) at an age that was more than 1 year younger than their current age. In these situations, the edited year- and month-of-last-use variables were assigned legitimate skip codes.

Prior substance use variables were not edited with respect to *imputed* core drug use variables. Suppose, for example, that respondents did not know or refused to report when they first used marijuana, but they gave ages of last use for marijuana in question LU02 that were consistent with their current ages. Although the potential existed for the imputed marijuana age at first use (AFU; imputed variable: IRMJAGE) to be imputed to a value greater than the age of last use, the marijuana age of last use (MRJAGLST) was not edited for consistency with IRMJAGE.

Similarly, if a core drug recency variable (e.g., MJREC for marijuana) had been assigned an "indefinite" value of 8 or 9 (indicating use at some point in the past 12 months or lifetime, respectively; see Section 6.2.2.3), the corresponding imputed recency (e.g., IRMJRC for marijuana) could be statistically imputed to indicate past month use (IRMJRC = 1). Although the

prior substance use module was designed for respondents who were not past month users, any data in this module were retained for respondents who were statistically imputed to be past month users. In this situation, analysts would have the option of deciding whether to use or disregard data from respondents who were imputed to be past month users.

Table B.19 describes the specific edits that were implemented for the age-, year-, and month-of-last-use variables in the prior substance use module (i.e., other than the edits described in Table B.18 when consistency checks were triggered between the age-, year-, and month-of-last-use variables). For example, a consistency check was triggered if respondents entered an age of last use that was earlier than the age when they reported first using drugs or cigarettes, or when they first smoked cigarettes daily. Because these age-of-last-use questions occurred in a noncore module, respondents were not allowed to change their answers to the corresponding age-at-first-use questions from the core modules. Thus, the only way that respondents could resolve the inconsistency between the age of last use and age at first use was to change their answer to the age-of-last-use question. If respondents indicated that their inconsistent age of last use was correct or they entered a new age-of-last-use value that was still inconsistent with the age at first use, the edited age-of-last-use variable (e.g., MRJAGLST for marijuana) was assigned a bad data code. Thus, the relevant age at first use from the core modules was used as the standard against which the corresponding noncore age-of-last-use variable was compared. Similarly, the year- and month-of-first-use questions for a given drug from the core section of the interview were used as standards for editing the year- and month-of-last-use variables for that drug.

As was the case in prior years, methamphetamine users who reported lifetime but not past month use of methamphetamine in the core stimulants module were routed into the questions for last use of methamphetamine. Since 2007, methamphetamine users identified in the special drugs module (see Section 6.2.6) also have been routed into these questions. The question wording for the age, year, and month when respondents last used methamphetamine differed depending on whether use was reported in the core stimulants module or in the noncore special drugs module. If respondents reported methamphetamine use in the core stimulants module, the wording of questions for the age, year, and month of last use asked about use of "Methamphetamine, Desoxyn, or Methedrine." For users identified through the special drugs module, the questions about the age, year, and month of last use simply referred to use of "Methamphetamine."

Because methamphetamine users who reported use in the core stimulants module were asked about prescription forms of the drug (e.g., Desoxyn®), the edits described in Table B.19 continued to hold for the age, year, and month of last use of prescription stimulants when methamphetamine users from the core stimulants module reported a later age, year, and month of last use than they reported for last use of *prescription* stimulants. However, these edits did not apply when respondents were identified as methamphetamine users through the special drugs module. In this latter situation, the question wording "Methamphetamine" was assumed to refer to the "street" form of the drug and not to prescription forms. For this reason, documentation of codes that indicated lifetime nonuse of stimulants was modified from "NEVER USED STIMULANTS" to "NEVER USED PRESCRIPTION STIMULANTS" beginning in 2007.

7.4.7.3 Sources of Psychotherapeutic Drugs

Since 2005, the prior substance use module has included questions on how nonmedical users of prescription pain relievers, prescription tranquilizers, prescription stimulants, methamphetamine, and prescription sedatives obtained the medications they misused in the past 30 days or past 12 months. For stimulants, these questions differentiated between stimulants that are typically available by prescription and methamphetamine, which is typically manufactured outside the legitimate pharmaceutical industry by illicit laboratories and distributed through illegal trafficking.[111]

For all of these drugs except methamphetamine, respondents were given a list of 10 potential sources of prescription medications. These sources included prescriptions from one or more doctors, fake prescriptions, thefts from medical facilities, the Internet, drug dealers, and friends or relatives (obtained with or without the knowledge of friends or relatives). Respondents also had the option of indicating that they obtained these medications "in some other way." Respondents who indicated that they obtained these medications in some other way were asked to specify what that other way was.

Questions about how methamphetamine users obtained this drug included a reduced list of six potential sources of the drug. Unlike the psychotherapeutic drugs that often may be available by prescription, the methamphetamine questions did not include options for respondents obtaining methamphetamine by prescription (including fake prescriptions written for methamphetamine) or by stealing methamphetamine from medical facilities or pharmacies. Response options for methamphetamine included obtaining the drug from a friend or relative for free, buying it from a friend or relative, taking it from a friend or relative without asking, buying it from a drug dealer or other stranger, buying it on the Internet, or getting it "in some other way." Again, respondents who reported that they got methamphetamine in some other way were asked to specify how they obtained it.

Since 2006, respondents who reported that they obtained pain relievers, tranquilizers, stimulants, sedatives, or methamphetamine from a friend or relative for free have been asked follow-up questions (e.g., LU27A for pain relievers) on how the friend or relative originally obtained the drug. These items were added in 2006 due to the large number of respondents in 2005 who reported obtaining these drugs from a friend or relative for free. Similar to their predecessor questions, for all of these drugs except methamphetamine, respondents were given the same list of 10 sources for where a friend or relative could have obtained the drugs. Respondents also had the option of indicating that the friend or relative obtained these medications "in some other way"; in that situation, respondents then were asked to specify the other way that the friend or relative obtained the medication.

[111] Although methamphetamine also is available in prescription form (e.g., Desoxyn®), legitimate prescribing of methamphetamine in the United States is relatively rare.

Respondents who reported that they last used a given psychotherapeutic drug in the past 30 days[112] were asked to report all of the ways that they obtained that drug in the past 30 days (e.g., question LU27 for pain relievers). Thus, the questions pertaining to how past month nonmedical users obtained these drugs in that period were "enter all that apply" questions in which respondents could indicate more than one source of these drugs. Each response option, such as "I got the pain reliever from a friend or relative for free" in question LU27 for pain relievers, was captured as a separate variable (ANLFRFRE for this option). The individual variables were coded as 1 if the response was chosen and were coded as 6 if the response was not chosen (Section 2.4.4). Codes of 94 and 97 (for "don't know" and "refused" respectively) in these variables indicated that respondents did not know or refused to report the source of the psychotherapeutic drugs that they used nonmedically in the past 30 days (Section 2.2.1).

Similar edits applied to the variables pertaining to how friends or relatives obtained medication that respondents subsequently used nonmedically. Specifically, if respondents reported that they last used a psychotherapeutic drug nonmedically in the past 30 days and they reported that they got the drug from a friend or relative for free, respondents were asked to report all of the ways that the friend or relative obtained that drug in the past 30 days (e.g., LU27A for pain relievers). In this situation, the corresponding variables were coded as 1, 6, 94, or 97, as described previously.

Respondents who used these drugs in the past 30 days and indicated more than one source of these drugs in that period were asked to report how they obtained the drugs that they last used (e.g., question LU28 for pain relievers). Similarly, respondents whose most recent use of a given psychotherapeutic drug was more than 30 days ago but within the past 12 months[113] were asked to report how they obtained the drug the last time that they used it. Unlike the "enter all that apply" variables pertaining to how past month nonmedical users obtained these drugs, respondents were allowed to enter only one response for how they obtained the psychotherapeutics for their last nonmedical use. Thus, for the edited variables ANLGTLAS, TRNGTLAS, STMGTLAS, and SEDGTLAS (pertaining to how respondents got pain relievers, tranquilizers, prescription stimulants, or sedatives, respectively, the last time they used them), codes in these variables corresponded to those in the corresponding questions LU28, LU30, LU32, and LU36 for these drugs.

If respondents reported that they last obtained a psychotherapeutic drug from a friend or relative for free in the past 12 months (e.g., in ANLGTLAS for pain relievers), similar edits applied to the variables pertaining to how friends or relatives obtained the medication that respondents last used nonmedically. Respondents could indicate only one way that the friend or relative obtained the psychotherapeutic drug that respondents last used nonmedically.

---

[112] For pain relievers, this included respondents who reported using OxyContin® nonmedically in the past 30 days. For methamphetamine, this included respondents who reported last using methamphetamine in the follow-up questions in the special drugs module and those who reported last using methamphetamine with a needle in the special drugs module (see Section 6.2.6), as well as those respondents who reported last using methamphetamine in the past 30 days in the core stimulants module.

[113] Again, for pain relievers, this included nonmedical use of OxyContin® more than 30 days ago but within the past 12 months. For methamphetamine, this included indications of use more than 30 days ago but within the past 12 months from either the core stimulants module or from questions in the noncore special drugs module (see Section 3.1).

For methamphetamine, question LU34 pertaining to how respondents obtained methamphetamine the last time they used it (and question LU34A pertaining to how friends or relatives obtained the methamphetamine that respondents last used) contained a reduced number of response options. Specifically, response options were not offered to respondents for obtaining methamphetamine via prescription (i.e., including fake prescriptions) or stealing it from medical facilities or pharmacies. Response categories in questions LU34 and LU34A were recoded in the corresponding edited variables MTHGTLAS (i.e., how the respondent obtained the methamphetamine the last time he or she used it) and MTHFFLAS (i.e., how the friend or relative obtained the methamphetamine that the respondent last used) to match the corresponding levels in the variables pertaining to the other psychotherapeutics. For example, level 5 in ANLGTLAS was "I got the pain reliever from a friend or relative for free." Similarly, level 5 in ANLFFLAS (i.e., how the friend or relative obtained the prescription pain reliever that the respondent last used nonmedically) was "He or she got the pain reliever from another friend or relative for free." In question LU34 for methamphetamine, however, "I got the Methamphetamine from a friend or relative for free" was the first response category. Similarly, the first response category in question LU34A was "He or she got the Methamphetamine from another friend or relative for free."

Therefore, if respondents reported that the last time they used methamphetamine, they got it from a friend or relative for free (LU34 = 1), that response was recoded to 5 in the edited variable MTHGTLAS. Similarly, if a respondent reported that he or she got the methamphetamine that he or she last used from a friend or relative for free, and the friend or relative had gotten that methamphetamine from another friend or relative for free (LU34A = 1), that response was recoded to 5 in the edited variable MTHFFLAS.

Consistent with overall editing procedures for the prior substance use module, an important aspect of editing the variables pertaining to sources of psychotherapeutic drugs involved assigning appropriate legitimate skip codes (Section 2.4.2). Conditions under which specific legitimate skip codes were assigned are discussed below for pain relievers, tranquilizers, and sedatives. Special issues for prescription stimulants and methamphetamine are discussed separately.

- If respondents reported in the relevant core section of the interview that they never were nonmedical users of prescription pain relievers, tranquilizers, or sedatives, the variables pertaining to how respondents obtained these drugs for nonmedical use and how friends or relatives obtained these drugs were assigned codes of 91 (e.g., NEVER USED PAIN RELIEVERS).
- If respondents were logically inferred in the relevant core section of the interview to have never used prescription pain relievers, tranquilizers, or sedatives nonmedically, the variables pertaining to how respondents obtained these drugs for nonmedical use and how friends or relatives obtained these drugs were assigned codes of 81 (e.g., NEVER USED PAIN RELIEVERS Logically assigned).
- If the variables for a given psychotherapeutic drug (e.g., pain relievers) had been skipped because respondents refused to answer all lifetime questions in the corresponding core module regarding whether they had ever used that type of drug

nonmedically, the refusal was propagated to the skipped source of psychotherapeutics variables, including the variables on how friends or relatives obtained them.

- If respondents reported that their last nonmedical use was more than 30 days ago but within the past 12 months, the variables pertaining to how respondents obtained that drug for nonmedical use in the past 30 days and how friends or relatives obtained it were assigned codes of 93 (e.g., DID NOT USE PAIN RELIEVERS IN THE PAST 30 DAYS).
- If respondents reported that their last nonmedical use was more than 12 months ago, the variables pertaining to how respondents obtained that drug for nonmedical use in the past 30 days and how friends or relatives obtained it were assigned codes of 93 (same meaning as above). In addition, the variables pertaining to how respondents obtained that drug the last time that respondents misused it in the past 12 months (e.g., ANLGTLAS and the associated "OTHER, Specify" variable ANLGTOSP for pain relievers) and how friends or relatives obtained the drug were assigned codes of 93 (e.g., DID NOT USE PAIN RELIEVERS IN THE PAST 12 MONTHS).
- If respondents reported that they got the drug from only one source in the past month (e.g., only one answer chosen in question LU27 for pain relievers), the corresponding variable for how respondents got the drug the last time they used it (e.g., ANLGTLAS for pain relievers) was assigned a legitimate skip code. The corresponding "OTHER, Specify" variable (e.g., ANLGTOSP for pain relievers) also was assigned a legitimate skip code. In these situations, it was not necessary to ask respondents how they got the drug the last time because they logically would have gotten it from that single source.
- If respondents used a drug nonmedically in the past 30 days but they did not report that one of the ways they obtained it was from a friend or relative for free, the corresponding edited variables pertaining to how friends or relatives obtained the drug were assigned legitimate skip codes.
- If respondents were asked how they obtained a given psychotherapeutic drug the last time they used it nonmedically and they did not report that they obtained it from a friend or relative for free, the corresponding edited variable pertaining to how the friend or relative obtained the drug was assigned a legitimate skip code.

As noted previously, questions in the prior substance use module distinguished between how respondents obtained *prescription* stimulants and how they obtained methamphetamine. The logic for assigning codes of 91 or 93 that was described previously in this section for pain relievers, tranquilizers, and sedatives also applied to the variables for the source of prescription stimulants if respondents reported in the core stimulants module that they never used methamphetamine or any prescription-type stimulants nonmedically, or if their most recent reported use of any stimulants from the core stimulants module caused them to be skipped out of the questions pertaining to how they obtained prescription stimulants in the past 30 days or the past 12 months. Similarly, if respondents reported that they obtained prescription stimulants from only one source in the past 30 days, the questions regarding how they obtained prescription stimulants the last time they used them nonmedically (i.e., STMGTLAS and STMGTOSP) were assigned legitimate skip codes.

In addition, respondents were not asked how they obtained prescription stimulants for nonmedical use if the only stimulant they reported ever using in the core stimulants module was methamphetamine. In this situation, the variables pertaining to the source of *prescription* stimulants were assigned legitimate skip codes. This edit also was implemented when respondents reported lifetime use of only two stimulants: methamphetamine and "some other stimulant," but the only "other" stimulant they specified using was methamphetamine. In this situation, data in the source of prescription stimulants variables were overwritten with codes of 89.

Although the core stimulants module did not explicitly ask when respondents last used stimulants other than methamphetamine, no editing was done to the prescription stimulants variables if the core recency-of-use variables for any stimulants (STIMREC) and methamphetamine (METHREC) indicated use in the same period, such as if STIMREC and METHREC both indicated use in the past 30 days. For example, if respondents were lifetime nonmedical users of methamphetamine and other stimulants and they indicated that they last used methamphetamine in the past 30 days, they also were expected to have reported that they used any stimulants in the past 30 days. Because questions LU31 and LU32 explicitly asked respondents how they obtained prescription stimulants, it was assumed when STIMREC and METHREC both indicated use in the same period that respondents' answers to questions LU31 and LU32 pertained to how they obtained prescription stimulants, and not how they obtained methamphetamine.

Because of the relationship between the core recency variables STIMREC and METHREC, however, a consistency check was triggered in the core stimulants module if respondents reported more recent use of methamphetamine (from the core question ST19) than they reported for any stimulants (from the core question ST09). In some of these situations, it was less clear whether respondents' answers in LU31 or LU32 referred specifically to prescription stimulants and not to methamphetamine. Therefore, a flag variable (STMGTFLG) was created when a consistency check was triggered between the recency of use for any stimulant and methamphetamine in the core stimulants module. The default value in STMGTFLG was 98 (BLANK [NO ANSWER]). Situations in which STMGTFLG had values other than 98 are described below.

STMGTFLG was set to a value of 1 if valid values existed in LU31 or LU32 for how respondents obtained prescription stimulants for nonmedical use and either of the following occurred:

- respondents answered the consistency check question STCC18 (which would indicate that their previous answer in the general stimulant recency question ST09 was incorrect),[114] regardless of whether they resolved the inconsistency between the most recent use of any stimulant and methamphetamine; or

[114] Question STCC18 is asked if respondents indicated in question STCC17 that their methamphetamine recency from question ST19 was correct (i.e., and by extension, that their general stimulant recency was incorrect) or that neither answer to their general stimulant recency and methamphetamine recency was correct. In question STCC18, respondents are asked again to report when they last used any stimulant nonmedically.

- respondents did not resolve the inconsistency, but the edited stimulant recency STIMREC was logically edited to more recent use based on the methamphetamine recency METHREC.

For example, suppose respondents reported in question ST09 that they last used any stimulant "more than 12 months ago," but they reported in ST19 that they last used methamphetamine "more than 30 days ago but within the past 12 months." If these respondents changed their stimulant recency to "more than 30 days ago but within the past 12 months" in STCC18, they would be routed to question LU32, regarding how they obtained *prescription* stimulants the last time they used them nonmedically. If these respondents reported obtaining prescription stimulants in a way that persons also might obtain methamphetamine, it could be questionable to assume that these answers in LU32 referred specifically to prescription stimulants and not to methamphetamine. STMGTFLG also could be set to a value of 1 if respondents reported in question STCC17 that "neither answer was correct" for their most recent use of any stimulant or methamphetamine and they indicated use of stimulants in the past month or past year in STCC18.

STMGTFLG was set to a value of 2 if valid values existed in LU31 or LU32 when the following occurred:

- respondents answered STCC19 (but not STCC18, which would indicate that their previous answer for when they last used methamphetamine was incorrect, but that the stimulant answer was correct), and
- they revised their methamphetamine recency (METHREC) to be consistent with the recency for any stimulants (STIMREC).

For these cases where STMGTFLG = 2, it could be more reasonable to assume that answers in LU31 or LU32 pertained to prescription stimulants and not methamphetamine. However, this value in STMGTFLG would still alert analysts to the occurrence of an inconsistency in the core stimulants data between when respondents reported last using any stimulants and methamphetamine.

The procedures for assigning values to STMGTFLG did not include situations in which LU31 or LU32 had missing values because the respondents had not resolved the inconsistency between the most recent use of any stimulant and methamphetamine, and STIMREC had been edited to infer use in the past month or past year, based on data in METHREC. Suppose, for example, that respondents initially reported last using methamphetamine in the past 30 days but they reported last using any stimulant more than 30 days ago but within the past 12 months, and they did not resolve this inconsistency when prompted to do so. For these respondents, STIMREC was assigned a value of 11 (Used in the past 30 days LOGICALLY ASSIGNED; see Sections 6.2.2.3 and 6.2.2.6). Because these respondents reported last using stimulants more than 30 days ago but within the past 12 months, they would be asked LU32 but they would be skipped out of LU31. In this example, the default assumption was applied that the answers in LU32 pertained to how respondents obtained prescription stimulants because respondents appeared to be making a distinction between "stimulants" and methamphetamine.

Also in this example, no editing was done to the variables from LU31 regarding how respondents obtained prescription stimulants for nonmedical use in the past 30 days. Specifically, it was not inferred that the source of prescription stimulants for respondents' last nonmedical use of prescription stimulants from LU32 applied to how respondents obtained prescription stimulants in the past 30 days because respondents may have used only methamphetamine but not prescription stimulants in the past 30 days. Likewise, values of 93 were not assigned to the 30-day prescription stimulant variables because some respondents may have misused prescription stimulants in the past month.

The skip logic for the variables pertaining to how respondents obtained methamphetamine took into account respondents' answers to the core methamphetamine questions in the stimulants module and the follow-up questions on methamphetamine from the special drugs module (see Section 6.2.6). Thus, codes of 91 (Section 2.4.2) were assigned to the methamphetamine variables if respondents reported one of the following:

- they indicated in both the core stimulants module and on follow-up in special drugs that they never used methamphetamine;
- they did not know or refused to report in the core stimulants module whether they ever used methamphetamine, but they indicated in special drugs that they never used it; or
- they explicitly indicated in the core stimulants module that they never used methamphetamine, but they did not know or refused to report on follow-up in the special drugs module whether they had ever used it.

Similarly, codes of 93 were assigned to the source of methamphetamine variables in one of two ways: (1) respondents reported in the core stimulants module that their last use of methamphetamine was outside of the period(s) of interest for asking the methamphetamine questions LU33 or LU34, or (2) respondents did not report methamphetamine use in the core stimulants module, but they reported use in special drugs, with their last use being outside of the period(s) of interest for LU33 or LU34. For example, if respondents did not report methamphetamine use in the core stimulants module but they reported in the special drugs question SD17b that they last used it more than 30 days ago but within the past 12 months, the variables corresponding to question LU33 were assigned codes of 93, and data from LU34 (and the "OTHER, Specify" variable LU34SP, if applicable) were assigned to the edited variable MTHGTLAS (and to MTHGTOSP, if applicable). Codes of 93 also were applied to the variables pertaining to how friends or relatives obtained methamphetamine if they did not use methamphetamine in the period of interest. For example, if respondents did not use methamphetamine in the past 12 months, then the edited variables MTHFFLAS and MTHFFLSP (pertaining to how friends or relatives obtained the methamphetamine that respondents last used) were assigned codes of 93.

Miscellaneous edits based on skip logic also applied to the source of psychotherapeutics variables. These edits applied to all of the psychotherapeutic drugs, with situations being cited for pain relievers. For example, if respondents used prescription pain relievers nonmedically in the past 30 days and they did not indicate that they obtained prescription pain relievers "in some other way" in that period in question LU27, then the corresponding "OTHER, Specify" variable

(e.g., ANLOTHSP, corresponding to question LU27SP) was assigned a code of 99. If respondents refused to report how they got pain relievers that they used nonmedically in the past 30 days, ANLOTHSP also was assigned a code of 97 (REFUSED). Similarly, if respondents used pain relievers nonmedically in the past 12 months, were asked question LU28, and did not indicate that they got pain relievers "in some other way" the last time they used them nonmedically, then the edited "OTHER, Specify" variable ANLGTOSP (corresponding to question LU28SP) was assigned a code of 99. If respondents refused to report how they got pain relievers the last time they used them nonmedically, ANLGTOSP also was assigned a refusal code.

Levels 1 through 9 in the "OTHER, Specify" variables were used for responses that corresponded to existing response options. For example, a code of 5 was assigned to ANLOTHSP (corresponding to question LU27SP) if respondents reported in LU27SP that they got pain relievers in the past 30 days from a friend or relative for free. These same coding categories applied to the "OTHER, Specify" variables for methamphetamine. Thus, a code of 5 also was assigned to MTHOTHSP (corresponding to question LU33SP) if respondents reported getting methamphetamine from a friend or relative for free in the past 30 days, even though this was the first response option in question LU33.

Table B.20 describes additional edits that were relevant to the source of psychotherapeutics variables. For these edits, the source of pain relievers variables are used as examples, although these edits also applied to the other psychotherapeutic drugs. Where relevant, the edits that are described in Table B.20 for how respondents obtained psychotherapeutic drugs also were applied to the variables for ways that friends or relatives obtained psychotherapeutic drugs.

7.4.7.4 Sequence of Initiation

If respondents first used alcohol and cigarettes, cigarettes and marijuana, alcohol and marijuana, or all three substances at the same age, they were asked to report which of these they used first. For example, if respondents indicated that they first used alcohol and marijuana at the same age, they were asked which of these they had used first.

- Questions LU22, LU23, and LU24 (corresponding to the edited variables USEALCG, USEMJCG, and USEALMJ, respectively) were asked when respondents reported first use of only two of these substances at the same age (i.e., USEALCG = use of alcohol and cigarettes at the same age; USEMJCG = use of marijuana and cigarettes at the same age; and USEALMJ = use of alcohol and marijuana at the same age).
- Questions LU25 and LU26 (corresponding to the edited variables USEACM and USENEXT) were asked when respondents reported first use of all three of these substances at the same age; USEACM indicated which of these three substances the respondents used first, and USENEXT indicated which of the remaining two substances the respondents used next.

As was the case for the processing of other variables in the prior substance use module, an important aspect of editing these sequence-of-use variables involved assigning various

legitimate skip codes, as appropriate. Conditions under which specific legitimate skip codes were assigned are discussed below.

- If respondents never used alcohol or cigarettes (regardless of whether they ever used marijuana), USEALCG was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED ALCOHOL/CIGARETTES.
- If respondents never used marijuana or cigarettes (regardless of whether they ever used alcohol), USEMJCG was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED MARIJUANA/CIGARETTES.
- If respondents never used alcohol or marijuana (regardless of whether they ever used cigarettes), USEALMJ was assigned a code of 91. Documentation of a code of 91 for this variable was 91 = NEVER USED ALCOHOL/MARIJUANA.
- If respondents never used alcohol, cigarettes, or marijuana, USEACM and USENEXT were assigned a code of 91. Documentation of a code of 91 for these variables was 91 = NEVER USED ALCOHOL/CIGARETTES/MARIJUANA.
- If the values in the edited age-at-first-use variables for alcohol (ALCTRY), cigarettes (CIGTRY), and marijuana (MJAGE) all were valid and equal, USEALCG, USEMJCG, and USEALMJ were assigned a code of 99 (LEGITIMATE SKIP).
- If at least one value for ALCTRY, CIGTRY, or MJAGE was valid but the values for all three were not equal, USEACM and USENEXT were assigned a code of 99.
- If at least one value for ALCTRY or CIGTRY was valid but the values were not equal, USEALCG was assigned a code of 99. Similarly, if at least one value for MJAGE or CIGTRY was valid but the values were not equal, USEMJCG was assigned a code of 99. If at least one value for ALCTRY and MJAGE was valid but the values were not equal, USEALMJ was assigned a code of 99.

Miscellaneous skip issues also applied to the data, such as if USEALCG (corresponding to LU22) was blank for some reason other than those mentioned above but some value other than 85, 91, 98, or 99 existed in questions LU23, LU24, or LU25. In these situations, a legitimate skip code was assigned to the relevant skipped variable (e.g., USEALCG). Similarly, if USEACM and USENEXT (corresponding to LU25 and LU26) had been skipped but data existed in LU22, LU23, or LU24, then USEACM and USENEXT were assigned legitimate skip codes. This logic covered residual situations in which variables might be skipped but data existed in one of the alternate variables.

Table B.21 describes additional edits pertaining to these sequence-of-use variables. For example, if ALCTRY, CIGTRY, and MJAGE all had codes of 997 (REFUSED) because respondents refused to answer the age-at-first-use questions for these drugs, or because respondents refused to answer the lifetime use question (for alcohol or marijuana), all blank values in USEALCG through USENEXT were replaced with the two-digit refusal code of 97 (Section 2.4.3). If for some reason respondents were routed into any of the questions corresponding to these variables when all of these ages at first use had a refusal code, nonblank values in USEALCG through USENEXT were overwritten with bad data codes.

7.4.8 Substance Treatment Module

The substance treatment module asked about receipt of treatment services for the use of alcohol or other drugs, not counting cigarettes. Questions about the receipt of treatment services included questions about receipt of treatment in the respondents' lifetime and in the past 12 months, specific locations where respondents received treatment in the past 12 months, emergency room visits in the past 12 months related to their use of specific drugs, whether they were still in treatment, the length of time since they were last in treatment (if they were not currently in treatment), specific questions about their last (or current) treatment episode, whether they were enrolled in treatment on October 1, 2012, and whether the only treatment they received in the past 12 months was detoxification.

Since 2007, respondents who reported methamphetamine use in the special drugs module have been eligible to be asked questions in the substance treatment module. However, this logic change affected only those respondents who reported methamphetamine use in the special drugs module and did not report use of alcohol or other drugs in the core modules; users of methamphetamine and other drugs would have been eligible to be asked the substance treatment questions based on their reported use of other drugs. However, no respondents in 2012 reported that they received substance treatment based solely on their reported methamphetamine use in the special drugs module.

In addition, respondents who reported methamphetamine use in the special drugs module were eligible to be asked question TX05 about visits to a hospital emergency room in the past 12 months to receive treatment for their use of cocaine, heroin, marijuana, PCP, LSD, or methamphetamine, provided that (1) they reported having received treatment in the past 12 months in question TX02, and (2) they reported in question TX03 that they received treatment in the past 12 months for their use only of drugs or their use of both alcohol and drugs. Again, all respondents who reported methamphetamine use in the special drugs module and reported receiving treatment in the past 12 months for their use of drugs (with or without receipt of treatment for alcohol use) either reported that they did not visit a hospital emergency room to receive treatment for their use of cocaine, heroin, marijuana, PCP, LSD, or methamphetamine, or they reported use of one of these other drugs in addition to their use of methamphetamine.

Questions also have been included since 2004 to capture information about respondents' life history of substance treatment. Respondents who had ever received treatment for their use of alcohol or other drugs but did not receive treatment in the past 12 months were asked questions regarding their receipt of treatment for their use of alcohol, other drugs, or both, depending on the substances they had reported using.[115] Where relevant, these respondents subsequently were asked to report the ages when they first received treatment for alcohol, drugs, or both. Respondents who reported that they had received treatment in the past 12 months were asked similar questions. Questions about treatment life history for respondents who reported that they had received treatment in the past 12 months also took into account their answers to question TX03, regarding whether they received treatment in the past 12 months for their use of alcohol, drugs, or both.

[115] Since 2007, reports of substance use have been based on reports in the core modules or reports of methamphetamine use from the special drugs module.

Since 2006, questions TX52 and TX53 have been included in the substance treatment module to ask whether respondents had attended a self-help group in the past 12 months (question TX52; edited variable TX12MSHG), and if so, whether this was for alcohol use, drug use, or both (TX53; edited variable TX12SGAD). These items were administered to respondents who previously indicated that they used alcohol or other drugs but they did not report receiving *treatment* through a self-help group in the past 12 months; use of the term "treatment" is inconsistent with the language used in the self-help/recovery community. Therefore, these items were added to assess whether asking respondents about treatment they received could result in underestimates of attendance at self-help group meetings.

Questions about the last or current treatment episode were asked principally of respondents who reported that they received treatment in the past 12 months (question TX02 answered as "yes"); the logic also routed respondents to the last or current treatment questions if they did not know or refused to report in question TX02 whether they had received treatment in the past 12 months. If respondents received treatment in the past 12 months (or answered question TX02 as "don't know" or "refused") and reported in question TX07 that they were currently in treatment,[116] subsequent questions asked about the main location where they were receiving treatment, specific drugs for which they were receiving treatment, the primary drug for which they were receiving treatment (if treatment for more than one drug was reported), the length of time that they had been in treatment thus far, and anticipated payment sources for their current treatment. If respondents were asked question TX07 and did not report currently being in treatment, these subsequent questions pertained to their last treatment episode, such as the duration of their last treatment and the payment sources for their last treatment. Respondents who did not report that they were currently in treatment also were asked about the outcome of their last treatment.

The substance treatment module also included questions about respondents' perceived need for treatment in the past 12 months if they never received treatment or did not report that they received treatment in the past 12 months. Questions about respondents' perceived need for treatment included questions about specific drugs for which respondents thought they needed treatment and whether they made specific efforts to receive treatment in the past 12 months. In addition, respondents who received treatment in the past 12 months but did not report that they were currently in treatment were asked whether they felt the need for *additional* treatment in the past 12 months. Those respondents who reported that they felt the need for additional treatment were asked about the specific drugs for which they needed additional treatment and whether they made specific efforts to receive additional treatment.

As noted previously, the substance treatment module was relevant only for respondents who reported some lifetime use of alcohol or other drugs, not counting cigarettes. Therefore, all of the edited treatment variables were assigned codes of 91 (i.e., NEVER USED ALCOHOL OR DRUGS) if respondents were skipped out of the entire substance treatment module because they never used alcohol, illicit drugs (including methamphetamine), or prescription-type psychotherapeutics for nonmedical reasons (i.e., pain relievers, tranquilizers, stimulants, or sedatives).

---

[116] Question TX07 asks, "Are you currently receiving treatment or counseling for your [TXFILL1]?" where [TXFILL1] could be replaced with "alcohol use," "drug use," or "alcohol or drug use."

In situations where respondents' only lifetime use of drugs involved use of OTC medications that were reported in one or more of the psychotherapeutics modules (i.e., and they did not report methamphetamine use in the special drugs module), codes of 81 were assigned to all of the edited substance treatment variables (i.e., NEVER USED ALCOHOL OR DRUGS Logically assigned). This was done to signify that these respondents were logically inferred to be lifetime nonusers of alcohol or drugs. This code of 81 also set these respondents apart from those whose original answers indicated that they had never used any of these drugs. These edits for assigning codes of 91 or 81 also applied to the self-help variables TX12MSHG and TX12SGAD.

7.4.8.1 Receipt of Substance Treatment Services

An important aspect of the processing of the substance treatment variables involved assignment of relevant legitimate skip codes when it could be determined unambiguously from respondents' answers that subsequent questions did not apply. In particular, respondents who were lifetime users of alcohol or at least one other drug were asked if they had ever received treatment for their alcohol or other drug use, not counting cigarettes. If respondents reported that they never received treatment (i.e., TXEVER = 2), the CAI program skipped them out of all remaining questions pertaining to the receipt of treatment services. Thus, if respondents clearly indicated that they never received treatment, the skipped treatment service variables were assigned legitimate skip codes (Section 2.4.2). As described in Section 2.4.3, when the treatment service questions were skipped because respondents refused to indicate whether they ever received treatment, the edited variables were assigned a refusal code; if treatment service questions were skipped because respondents did not know whether they ever received treatment, the edited variables retained a value of blank.

Similarly, respondents were not asked subsequent questions about receipt of treatment services in the past 12 months if they did not report having ever received treatment in that period (i.e., TXYREVER = 2). Thus, if respondents reported that they did not receive treatment in the past 12 months and there were no other responses in the substance treatment module to suggest that they had (see below), legitimate skip codes were assigned to the variables pertaining to receipt of treatment in specific locations in the past 12 months. The procedures for editing 12-month treatment variables that had been skipped when respondents refused to indicate whether they had received treatment in the past 12 months or did not know whether they had received treatment in this period were the same as those described above.

If respondents reported that they received treatment in the past 12 months, it was possible for them to be asked subsequent questions about treatment in an emergency room in the past 12 months for their use of marijuana, cocaine, heroin, LSD, PCP, or methamphetamine. Respondents were not asked these questions if they previously reported that their treatment in the past 12 months was only for their use of alcohol. Thus, "legitimate skip" codes were assigned to the edited variables pertaining to emergency room use (TXYRVSER and TXYRNMER), provided there were no other answers in the substance treatment module that indicated treatment for use of these drugs, which would suggest that respondents should have been asked these questions. Similarly, legitimate skip codes were assigned to the edited variable pertaining to the number of emergency room episodes for treatment of these six drugs (TXYRNMER) if respondents reported that they never received treatment in an emergency room related to their use of these drugs.

In addition, respondents who reported receiving treatment in the past year were not asked certain questions about receipt of treatment related to their use of specific drugs if they were lifetime nonusers of these drugs. For example, respondents who never used heroin were not asked whether they last received (or were currently receiving) treatment for their use of heroin. Similarly, respondents who reported receiving treatment in the past 12 months but who never used marijuana, cocaine, heroin, LSD, PCP, or methamphetamine were not asked the questions about use of hospital emergency room services for the use of these drugs. Rather than assign the usual type of legitimate skip code (i.e., 99 or 89), however, a special code of 6 was assigned in these situations, provided that the respondent had not indicated receipt of treatment for any of these drugs elsewhere in the substance treatment module. This code had the following meaning: 6 = Never used the relevant drug.

This coding was done because respondents could be routed into or skipped out of a number of different combinations of questions depending on their reported drug use history. For example, a respondent who reported that he or she had received treatment in the past 12 months and was a lifetime user of alcohol, marijuana, cocaine, hallucinogens, prescription pain relievers, and prescription stimulants would selectively be asked the questions about treatment for these drugs during his or her last treatment or current episode and would not be asked the questions pertaining to treatment for heroin, inhalants, prescription tranquilizers, and prescription sedatives.

When respondents were skipped out of a question related to treatment for a given drug because they refused to indicate whether they had ever used that drug, the refusal was propagated onto the edited variable pertaining to treatment for that drug. For example, if a respondent reported receiving treatment in his or her lifetime but refused to indicate whether he or she had ever used heroin, the question about treatment for heroin during the last treatment episode was skipped. The edited variable pertaining to treatment for heroin (TXLTYHER) was therefore assigned a refusal code.

As noted above, respondents who did not report that they received treatment in the past 12 months were not asked questions about their last treatment episode. Therefore, if the final edited variable pertaining to receipt of treatment in the past 12 months indicated that respondents had not received treatment during this period (i.e., TXYREVER = 2), the variables pertaining to the last treatment episode were assigned legitimate skip codes.

Most of the editing of the substance treatment questions TX45 through TX51A that have been present since 2004 also involved assigning legitimate skip codes where relevant. Consistent with the logic described above, if respondents had never used alcohol or other drugs, these variables were assigned codes of 91 or 991 (or 81 or 981, if their only use of drugs involved OTC medications). In addition, if respondents reported in question TX01 that they had never received treatment, these variables were assigned legitimate skip codes (e.g., 99 or 999). Questions TX45 through TX48A also pertained to respondents who had received treatment but not in the past 12 months. Therefore, if respondents reported in question TX02 that they received treatment in the past 12 months, the edited variables corresponding to questions TX45 through TX48A were assigned legitimate skip codes. In addition, when respondents reported receiving treatment in the past 12 months, subsets of the variables corresponding to questions TX49 through TX51A were assigned legitimate skip codes based on answers in question TX03

regarding receipt of treatment in the past 12 months for alcohol, drugs, or both. Variables corresponding to TX49 through TX51A also were assigned legitimate skip codes based on indications in the core modules or in the special drugs module for methamphetamine that respondents never used alcohol or never used any illicit drugs. Similar assignment of legitimate skip codes occurred for the variables corresponding to TX49 through TX51A if respondents received treatment in their lifetime but not in the past 12 months, or depending on respondents' answers in the core drug modules or in special drugs for methamphetamine. For example, if respondents answered question TX02 as "no" regarding receipt of treatment in the past 12 months, the variables corresponding to TX49 through TX51A were assigned legitimate skip codes; by definition, these respondents had to have answered the lifetime treatment question TX01 as "yes."

Table B.22 in Appendix B presents additional edits that were specific to the variables for the receipt of treatment services for variables that existed prior to 2004. For example, the answers to the questions on receipt of treatment in the past 12 months and the last time that respondents received treatment could be inconsistent. Specifically, respondents could report that they received treatment in the past 12 months (TX02 = 1) but then subsequently report that the last time they received treatment was more than 12 months ago (TX24 = 3). For these respondents, the recency of treatment was inferred to be at some point within the past 12 months (TXLASREC = 8). Respondents also could provide an answer other than "yes" when asked in question TX02 whether they had received treatment in the past 12 months and then indicate that they last received treatment in the past 30 days or more than 30 days ago but within the past 12 months (TX24 = 1 or 2). In these situations, the respondents were logically inferred to have received treatment in the past 12 months. Similarly, respondents could answer "don't know" or "refused" when asked whether they had received treatment in the past 12 months and then report that they last received treatment more than 12 months ago. In this situation, a negative response was logically inferred for the variable pertaining to receipt of treatment services in the past 12 months (TXYREVER = 4).

In addition, composite variables combining data from more than one individual item were created for the following:

- the main place where respondents received (or were receiving) treatment during their last (or current) treatment episode (TXLTYMN);
- the outcome of the last treatment episode, for respondents who were not currently in treatment (TXLTYOUT); and
- the length of time that respondents had been in treatment or currently had been in treatment thus far (TXLTYDUR).

For the first two variables listed above, respondents could select a response category from a list, including selection of an "other" category (e.g., treatment in some other place). Only those respondents who chose the other category were routed into a second item where they were asked to specify the other location or the other outcome of their treatment. Consequently, the final variables for the main place where respondents received (or were receiving) treatment during their last (or current) treatment episode and the outcome for that last episode included data both from the existing response categories that respondents were allowed to choose and valid "other"

responses that they specified. If respondents chose the other category but specified something that was coded with a missing value (i.e., "bad data," "don't know," "refused," or blank), a final code of "other" was retained for these two variables.

The variable pertaining to the length of time that respondents had been in treatment (TXLTYDUR) was derived from a question that asked respondents to indicate whether they wanted to give their answer in terms of days, months, or years, and from questions that asked for the number of days, months, or years that they were in treatment. TXLTYDUR was expressed as the number of days that respondents were in treatment. If respondents answered in terms of a number of months, their reported number of months was multiplied by 30. If respondents answered in terms of a number of years that they had been in treatment, their reported number of years was multiplied by 365.

If respondents answered in terms of a number of months in treatment, the treatment duration data also were compared for consistency with the respondent's age. Specifically, the number of months in treatment was divided by 12 to yield an estimated number of years in treatment. If the reported number of years in treatment exceeded the respondent's current age, then TXLTYDUR was assigned a bad data code. If the difference between the respondent's current age and the number of years in treatment was 10 or fewer years, this data pattern was flagged. Such respondents would have been reporting that they had *not* been in treatment for 10 or fewer years. However, TXLTYDUR was not set to bad data for this latter situation.

Table B.23 presents edits that were specific to the substance treatment variables that were added to the survey in 2004. For example, respondents could report that they first received treatment for their use of alcohol at ages that were earlier than when they first reported using alcohol. No editing was done to these data. However, flags were created to indicate whether the ages for first treatment of alcohol or other drugs were consistent with reported ages at first use from the core modules or in the special drugs module for methamphetamine (since 2007), and if not, the flags indicated the degree of inconsistency between these data. The meaning of the values in these flag variables is discussed in more detail in Table B.23.

In addition, data from substance treatment variables that existed prior to 2004 were used to edit these added variables. However, data from these added substance treatment variables were not used to edit the substance treatment variables that existed prior to 2004. Consequently, variables that also existed prior to 2004 (as well as in 2004 through 2012) were created in a manner that was comparable with how these variables were created in prior years.

An important aspect of editing the variables TX12MSHG (corresponding to question TX52) and TX12SGAD (corresponding to TX53) involved assigning relevant codes based on the skip logic for these items. For example, if respondents were routed to question TX52 and answered it as "no," then TX12SGAD was assigned a legitimate skip code. Similarly, if TX52 was refused, that refusal was propagated to TX12SGAD. In addition, if the lifetime treatment variable TXEVER (corresponding to question TX01) had been set to bad data, then any nonblank values also were replaced with codes for bad data in TX12MSHG and TX12SGAD (see Table B.22 in Appendix B).

If question TX52 had been skipped because respondents had reported earlier in the substance treatment module that they received treatment in a self-help group in the past 12 months (from question TX04H), then TX12MSHG was assigned a code of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). In addition, TX53 was skipped when question TX04H was answered as "yes." Therefore, if values of 1, 2, or 3 existed in TXYRSGAD (corresponding to question TX04H1 and indicating treatment in a self-help group in the past 12 months for alcohol only, drugs only, or both alcohol and drugs, respectively), TX12SGAD (corresponding to TX53) was assigned corresponding codes of 11, 12, or 13. "Bumping" these values by 10 in the assignment of codes from TXYRSGAD to TX12SGAD indicated to analysts that a logical inference was made in TX12SGAD; in contrast, codes of 1, 2, or 3 in TX12SGAD came from respondents answering TX53 directly. Other values greater than 3 that existed in TXYRSGAD were moved over to TX12SGAD without further editing. In particular, if TXYRSGAD (from question TX04H1) had been legitimately skipped because respondents reported in question TX03 that they were treated in the past 12 months for alcohol only or drugs only, then a legitimate skip code could be assigned to TX12SGAD as well.

Variables pertaining to the lifetime receipt of treatment (TXEVER), treatment in the past 12 months (TXYREVER), treatment in a self-help group in the past 12 months (TXYRSHG), and treatment for alcohol, drugs, or both in a self-help group in the past 12 months (TXYRSGAD) were not edited to make them consistent with data in questions TX52 and TX53. Consequently, the integrity of trends in the receipt of treatment (particularly in the past 12 months) can be preserved. This decision also was in agreement with the decision that was made in 2004 not to edit substance treatment variables that existed in the module prior to 2004 for consistency with data from questions TX45 through TX51A that were added to the substance treatment module in 2004.

However, a flag variable (TXSHGFLG) was created to alert analysts to situations in which data were inconsistent between TX12MSHG and TXYRSHG. The default value of TXSHGFLG was 98 (blank). In addition, TXSHGFLG was assigned values of 91 if respondents had never used alcohol or other drugs, and it was assigned values of 81 if the logical inference had been made that respondents had never used alcohol or other drugs. Assignment of values other than 81, 91, or 98 to TXSHGFLG is described below.

- TXSHGFLG was set to a value of 1 if respondents answered TX52 as "yes" (i.e., attended a self-help group for their own use of alcohol or other drugs in the past year) but the respondents answered question TX04H as "no" (i.e., did not receive treatment in a self-help group in the past 12 months). In this situation, the answers in TX04H and TX52 may be seen as contradictory, except that respondents may not have thought of self-help group attendance as "treatment" when they were answering question TX04H.
- TXSHGFLG was set to a value of 2 if respondents answered TX52 as "yes" but they reported either that they did not receive any treatment in the past 12 months or in their lifetime. A separate level was created for this pattern because these respondents may not have thought of self-help groups as "treatment" when they were answering question TX01 (lifetime receipt of treatment) or TX02 (receipt of treatment in the

past 12 months). Consequently, respondents who attended self-help groups in the past 12 months may have answered TX01 or TX02 as "no."

- TXSHGFLG was set to a value of 3 if the edited variable TXYRSHG (corresponding to question TX04H) indicated that respondents were logically inferred to have received treatment in a self-help group in the past year, but TX12MSHG also did not indicate that these respondents attended a self-help group in the past 12 months. In this situation, TX12MSHG was inconsistent with TXYRSHG, but only because a logical inference had been made that TXYRSHG should have been answered as "yes."
- TXSHGFLG was set to a value of 4 if TX52 was answered as "yes" but the edited variable TXYRSHG had a value for "don't know," "refused," or "blank." In particular, TXYRSHG could have been blank if question TX04H had been skipped and it had subsequently been inferred that respondents got treatment in the past 12 months. In these types of situations, data in TX12MSHG and TXYRSHG were not necessarily contradictory.
- TXSHGFLG was set to a value of 5 if question TX04H was answered as "no" and TX52 was answered as "don't know" or "refused." Although a logical inference was not made in TX12MSHG that respondents did not attend self-help groups in the past 12 months, analysts would have the option of inferring that respondents absolutely did not attend self-help groups in the past 12 months for their own use of alcohol or other drugs based on TXYRSHG. Alternatively, analysts could decide to treat self-help group attendance data as missing based on TX12MSHG.

7.4.8.2 Perceived Need for Substance Treatment

The basic content of the section of the substance treatment module in 2012 pertaining to respondents' perceived need for substance abuse treatment did not change relative to 2011. Since 2007, however, respondents who reported methamphetamine use in the noncore special drugs module and who did not report that they received treatment for their alcohol or other drug use in the past 12 months have been eligible to be asked whether they felt that they needed treatment for their alcohol or other drug use at any time in the past 12 months. As noted previously, this logic change was relevant only for those respondents who reported methamphetamine use in the special drugs module and did not report use of alcohol or other drugs in the core modules. However, no respondents whose only substance use came from their report of methamphetamine use in the special drugs module and who reported that they never received substance treatment reported that they felt that they needed treatment in the past 12 months.

Since 2002, respondents have been asked to report the reasons why they did not receive substance treatment services despite feeling the need for treatment (question TX22A). Similarly, respondents who reported that they received treatment but needed additional treatment have been asked why they did not receive additional treatment or counseling (question TX23A).

Based on a review of what respondents had specified in quarter 1 of 2002 as leading "other" reasons for not getting substance treatment or additional treatment, additional questions (TX22B and TX23B) have been included in the interview since 2003 to capture commonly endorsed other reasons for not receiving treatment. Therefore, if respondents answered question

TX08 as "yes" (i.e., perceived the need for substance treatment), they eventually were routed to question TX22A, where they could report up to 10 reasons for not receiving substance treatment, including "some other reason or reasons." If respondents chose the "some other reason or reasons" option in question TX22A, they were routed next to question TX22B, where they could choose additional reasons for not getting substance treatment; again, respondents were given the option in question TX22B to report "some other reason or reasons." If respondents chose this "other" response category in TX22B, they were asked to specify the *most important* other reason why they did not get treatment. Thus, respondents were asked to specify only the most important other reason for not getting substance treatment. Similar logic was in place for questions TX23A and TX23B, when respondents reported in TX09 that they perceived a need for additional substance treatment.

Questions TX22A, TX22B, TX23A, and TX23B were "enter all that apply" questions in which respondents could choose more than one reason from each list. Each response option (e.g., "You had no health care coverage, and you couldn't afford the cost") was captured as a separate variable. The edited variables corresponding to the individual response options in TX22A, TX22B, TX23A, and TX23B were coded as 1 (Response entered) or 6 (Response not entered), if at least one item was chosen from the TX22A, TX22B, TX23A, and TX23B lists (see Section 2.4.4).

In addition, the 10th response option in questions TX22A and TX23A ("some other reason or reasons") was principally considered to be a "toggle" to questions TX22B and TX23B, respectively. Therefore, separate "some other reason or reasons" variables were not created to correspond to the last response category in TX22A and TX22B. Similarly, separate variables were not created to correspond to the last category in TX23A and TX23B.

The edits described below were implemented in 2003 to take into account the addition of questions TX22B and TX23B. The edits described use TX22A and TX22B as examples but also applied to TX23A and TX23B.

- If respondents chose the 10th response option in TX22A, any response that was entered from the TX22B series was coded as 1, and anything that respondents did not choose from the TX22B list was coded as 6.
- If respondents chose a response from TX22A but did not choose the 10th response category in TX22A, the variables corresponding to the response categories in TX22B (i.e., NDTXNOND through NDTXSOR) all were given a code of 6 (Response not entered), rather than being assigned "legitimate skip" codes. That is, TX22A and TX22B were considered together to be one big series of reasons.
- If respondents chose category 10 in question TX22A, chose at least one reason from TX22B, but did not choose category 6 in TX22B (some other reason or reasons), the edited variable NDTXSOR (some other reason or reasons for not receiving substance treatment) was assigned a code of 6. That is, it was inferred in this situation that the list of specific reasons in TX22B was adequate for capturing why respondents did not get substance treatment. For example, if a respondent chose category 10 in TX22A and then chose only category 2 in TX22B ("You thought you could handle the

problem without treatment"), it would be reasonable to infer that this response in TX22B was the only other reason why the respondent did not get treatment.

- If respondents chose response category 10 in question TX22A, it was possible for them to answer TX22B as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable NDTXSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reason or reasons" somewhere in the series. Remaining variables corresponding to the TX22B series retained codes of 94 ("don't know") or 97 ("refused").
- If respondents answered question TX22A as "don't know" or "refused," question TX22B was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the TX22B list.
- If NDTXSOR had a value of 6 (see above), the "OTHER, Specify" variable NDTXRIMP was assigned a legitimate skip code. If NDTXSOR had a refusal code, that refusal was propagated onto NDTXRIMP.
- If NDTXSOR had a code of 1 when the respondent answered "don't know" or "refused" to the TX22B series, the "OTHER, Specify" variable NDTXRIMP retained a code of 98 (blank).

Consistent with general editing procedures, if respondents reported a reason that corresponded to a reason in the lists for TX22A/TX22B or TX23A/TX23B, that reason was logically inferred to have been chosen in the relevant edited variable (see Section 2.4.5). Suppose, for example, that a respondent had not received treatment but felt the need for it, and the respondent specified that one of the reasons for not receiving treatment was that he or she was not ready to stop using alcohol or drugs. If the respondent had not chosen this response in TX22A, the edited variable NDTXREDY (corresponding to response category 5 in TX22A) was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). Similarly, if the respondents specified that they did not get treatment because they thought they could handle the problem without treatment but had not chosen that reason in TX22B, the edited variable NDTXHNDL (no substance treatment because the respondent thought he or she could handle the problem without treatment) was assigned a code of 3.

Conversely, if respondents did not report "some other reason" why they did not receive treatment in the past 12 months (edited variable NDTXSOR = 6, corresponding either to response category 10 in question TX22A not being chosen or response category 6 in question TX22B not being chosen), legitimate skip codes were assigned to the edited "OTHER, Specify" variable NDTXRIMP (corresponding to question TX22SP). Similar edits were done for the "OTHER, Specify" variable pertaining to reasons for not receiving additional treatment if respondents reported that they felt the need for additional treatment but did not indicate "some other reason" for not receiving additional treatment.

As was the case with the variables pertaining to receipt of treatment services, an important aspect of the processing of the variables pertaining to perceived need for treatment involved assigning relevant legitimate skip codes (Section 2.4.2). In particular, the variables on perceived need for treatment were compared with data on receipt of treatment services in the past

12 months. For example, if respondents had received treatment services in the past 12 months, the questions about perceived need for treatment in that period did not apply. Thus, legitimate skip codes were assigned to the variables pertaining to the perceived need for any alcohol or other drug treatment when respondents had received treatment in the past 12 months. Similarly, if respondents received treatment in the past 12 months and they reported that they were still in treatment (TXRCVNOW = 1), the questions about perceived need for additional services did not apply, and legitimate skip codes were assigned to the corresponding edited variables.

Respondents who had not indicated that they received treatment in the past 12 months and who were lifetime users of alcohol or some other drug also were skipped out of questions regarding their perceived need for additional treatment. Again, the edited variables corresponding to perceived need for additional services were assigned legitimate skip codes. Those respondents who had not indicated that they received treatment in the past 12 months were asked the general question about whether they perceived themselves as needing treatment for their use of alcohol or other drugs (edited variable NDTXNEDR). If they did not see themselves as needing treatment, they were skipped out of questions pertaining to perceived need for treatment for specific drugs in the past 12 months. Again, legitimate skip codes were assigned to the edited variables that had been skipped.

Similarly, respondents were globally skipped out of questions TX11 through TX22 (regarding their perceived need for any treatment for alcohol or specific other drugs) if they reported in question TX02 that they received treatment in the past 12 months. Therefore, the edited variables corresponding to questions TX11 through TX22 (NDTXALCR through NDTXEFTR) were assigned legitimate skip codes.

Legitimate skip codes also were assigned in situations in which respondents were lifetime nonusers of a particular drug. For example, if respondents indicated that they needed treatment for their use of alcohol or drugs, they were asked about their perceived need for treatment only for those specific drugs that they had ever used; legitimate skip codes were assigned to the skipped drug-specific variables that respondents had never used. Thus, for example, if a respondent had never used heroin but reported needing treatment in the past 12 months for alcohol or drugs (TX08 = 1), a legitimate skip code was assigned to the edited variable pertaining to the perceived need for treatment for heroin (NDTXHERR).

Procedures consistent with those described in Section 2.4.3 also were implemented when questions about the perceived need for treatment were potentially applicable, but respondents refused to report whether they had ever used a particular drug. For example, if a respondent had not received treatment in the past 12 months, reported needing treatment in the past 12 months for alcohol or other drugs, but refused to report whether he or she had ever used heroin, the item about perceived need for treatment for heroin was skipped. Because the respondent refused to report about lifetime use or nonuse of heroin, the edited variable NDTXHERR was assigned a refusal code.

Table B.24 presents additional edits that were specific to the variables pertaining to the perceived need for treatment services. As noted above, for example, respondents were skipped out of questions TX11 through TX22 if they reported that they received treatment in the past 12 months. If respondents had not originally reported receiving treatment in the past 12 months but

were logically inferred to have done so (see Table B.22), these respondents would have been routed to questions TX11 through TX22. Rather than replace respondents' answers with codes for "bad data," however, special codes were assigned to indicate that respondents were routed into questions about their perceived need for treatment for use of specific drugs when they were logically inferred to have received treatment in the past 12 months. This procedure would allow analysts to decide whether to use or disregard these data in their analyses.

7.4.9 Health Care Module

The health care module included questions for female respondents aged 12 to 44 regarding whether they were currently pregnant, and if so, the number of months that they had been pregnant. This section also included questions for all respondents regarding utilization of hospital emergency room services and overnight inpatient hospitalizations in the past 12 months, as well as lifetime and past year histories of specific health conditions.

An important aspect of processing the variables in this section involved assignment of legitimate skip codes, where relevant (Section 2.4.2). For example, males of any age and women over the age of 44 were assigned legitimate skip codes to the pregnancy variables. Similarly, if females aged 12 to 44 reported that they were not currently pregnant (PREGNANT = 2), legitimate skip codes were assigned to the variable pertaining to the number of months that they were pregnant (PREGMOS).

In the pregnancy variables, if women reported currently being pregnant, the allowable range for the number of months that they were pregnant ranged from 1 to 9 months. Thus, women who reported that they were currently pregnant were not allowed to report that they had been pregnant for "0" months.

In the health care questions, respondents who did not report that they were hospitalized overnight in the past 12 months (edited variable INHOSPYR) were not asked for the number of times they were hospitalized in that period (edited variable NMNGTHSP). If respondents reported that they were not hospitalized overnight in the past 12 months (INHOSPYR = 2), the variable NMNGTHSP was assigned a legitimate skip code. If respondents refused to report whether they were hospitalized overnight in the past 12 months (INHOSPYR = 97), that refusal was propagated onto NMNGTHSP (Section 2.4.3).

The allowable range for the question about the number of nights that respondents were inpatients in a hospital in the past 12 months included 365. No editing was done to the variable NMNGTHSP when respondents reported that they had spent all 365 nights in a hospital in the past 12 months.

Questions have been included in the health care module since 2005 about the occurrence of the following health conditions in the lifetime and past 12 month periods: anxiety disorder, asthma, bronchitis, cirrhosis of the liver, depression, diabetes, heart disease, hepatitis, high blood pressure, HIV/AIDS (i.e., human immunodeficiency virus or acquired immunodeficiency syndrome), lung cancer, pancreatitis, sexually transmitted disease (STD, such as chlamydia, gonorrhea, herpes, or syphilis), sinusitis, sleep apnea, stroke, tinnitus, tuberculosis, or ulcers. The content of these items and associated edits have not changed.

Respondents were asked whether a doctor or other medical professional had ever told them that they had any of the specific health conditions mentioned above. Respondents could report that they had been told that they had as many of these conditions as applied (i.e., this was an "enter all that apply" question). Respondents also could report that they never had any of these conditions. However, if they reported having one of the conditions listed above and also that they never had any of these conditions, the CAI program triggered an error message that required the respondents to resolve the inconsistency before they could proceed further.

If respondents reported that they had been told that they ever had some of these specific conditions, they were asked whether a doctor or other medical professional told them they had these specific conditions in the past 12 months. The CAI logic restricted respondents' choices in the past 12 months to those conditions that they reported for the lifetime period. For example, if a respondent reported ever being told by a doctor or other health professional that he or she had asthma and bronchitis, but the respondent did not indicate being told that he or she had any of the other health conditions, the respondent's choices for the past 12 month period were limited to reporting whether a doctor or health professional told the respondent that he or she had asthma, bronchitis, or none of these conditions; if the respondent attempted to choose another response for a condition in the past 12 months (e.g., diabetes), the CAI program triggered an error message that this was not one of the respondent's choices. Similarly, respondents were not allowed to report that they had one or more health conditions in the past 12 months and that they had "none of the above" (i.e., none of these conditions in the past 12 months).

Because these were "enter all that apply" variables, separate variables were created for each health condition for the lifetime and past year periods (e.g., LIFANXD and LIFASMA for lifetime occurrence of anxiety disorders or asthma, respectively; YRANXD and YRASMA for the occurrence of these respective conditions in the past year). The individual edited variables for these lifetime and past year health conditions were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (see Section 2.4.4).

The CAI logic discussed above eliminated the occurrence of the inconsistent data patterns noted above for these health condition variables, namely reports of having specific health conditions and having none of them, or respondents not reporting these conditions for the lifetime period but reporting them for the past 12 months. Consequently, the editing procedures for these health condition variables involved assignment of legitimate skip codes based on the skip/routing logic. These edits are discussed below.

- If respondents reported that they never had any of these conditions in the lifetime period (i.e., edited variable LIFNONE coded as 95), all of the variables pertaining to lifetime medical conditions (LIFANXD through LIFULCER) were assigned legitimate skip codes. In addition, all of the past year variables (YRANXD through YRULCER and also YRNONE) were assigned legitimate skip codes.
- If respondents reported the lifetime occurrence of at least one of these conditions, LIFNONE was assigned a legitimate skip code.
- If respondents reported the lifetime occurrence of at least one of these conditions but that a doctor or other health professional did not tell them that they had any of these conditions in the past 12 months (i.e., edited variable YRNONE coded as 95), all of

the variables pertaining to past year medical conditions (YRANXD through YRULCER) were assigned legitimate skip codes.

- If respondents reported the lifetime occurrence of at least one of these conditions and they did not report the lifetime occurrence of specific other conditions, the corresponding past year variables for the conditions they did not report were assigned legitimate skip codes. For example, if respondents reported that a doctor or other health professional had ever told the respondents that they had diabetes (LIFDIAB = 1) but they did not report ever being told that they had high blood pressure (LIFHBP = 6), the past year high blood pressure variable YRHBP was assigned a legitimate skip code.
- If respondents reported the lifetime occurrence of a particular condition but they did not report that a doctor or other health professional told them that they had this condition in the past year, the edited variable for the condition in the past year was coded as 6 (Response not entered). Suppose, for example, that a respondent reported the lifetime occurrence of bronchitis and high blood pressure (LIFBRONC = 1 and LIFHBP = 1, respectively), and the respondent reported being told in the past year that he or she had high blood pressure but the respondent did not report being told in the past year that he or she had bronchitis, then YRBRONC was coded as 6 and LIFHBP was coded as 1 (Response entered).
- If respondents reported that a doctor or other health professional told them in the past year that they had one or more specific health conditions, YRNONE was assigned a legitimate skip code.

Because these health condition questions were "enter all that apply" items, if respondents indicated that they did not know or refused to report whether they had any of these conditions in the lifetime or past year periods, codes of 94 (for "don't know") or 97 (for "refused") were propagated to all of the unedited variables corresponding to these health conditions in the relevant time period, including the unedited variables indicating that respondents had none of these conditions. Therefore, if respondents reported at least one lifetime health condition but they answered the past year question as "don't know" or "refused," the corresponding codes of 94 or 97 that were assigned to the past year health conditions that the respondents did not report having in their lifetime were replaced with codes of 89 (LEGITIMATE SKIP Logically assigned). In this situation, YRNONE retained the code of 94 or 97 because respondents may not have been told in the past year that they had any of the conditions that they reported for the lifetime period. These edits preserved those responses of "don't know" or "refused" in the past year variables that correspond to conditions that respondents reported that they had in their lifetime.

For example, suppose a respondent reported being told that he or she had diabetes in his or her lifetime but did not choose the lifetime item for lung cancer (i.e., LIFDIAB = 1 but LIFLUNCA = 6). If the respondent answered the past year question as "don't know," the edited past year variable for diabetes (YRDIAB) retained a code of 94 but the past year variable for lung cancer (YRLUNCA) was assigned a code of 89. In addition, YRNONE retained a code of 94.

In addition, if respondents answered the question for the lifetime list of health conditions as "don't know" or "refused," the corresponding code of 94 or 97 was propagated to LIFNONE

as well. The item also was skipped pertaining to the occurrence of these conditions in the past year. In this situation, the code of 94 or 97 was retained in LIFNONE because respondents may never have been told by a health professional that they had any of these conditions. The relevant code of 94 or 97 also was propagated to the individual past year variables. For example, if respondents did not know whether a health professional had ever told them that they had any of these conditions, codes of 94 were assigned to the edited past year variables YRANXD through YRULCER and to YRNONE. That is, if these respondents did not know whether they had ever been told that they had any of these conditions, it could reasonably be inferred that the respondents did not know whether they had any of them in the past year. This edit served to reduce the number of codes of "blank" in the corresponding past year variables.

No editing was done in situations in which respondents reported that they had been told at some point in their lifetime that they had certain long-term chronic medical conditions (e.g., cirrhosis of the liver, HIV/AIDS) but did not report these conditions for the past year period. The rationale for not editing the data in this situation was that the past year question asked respondents to indicate which of these conditions a doctor or medical professional told them that they had in the past 12 months. Consequently, respondents may have had these chronic conditions in the past 12 months, but a doctor or other health professional literally may not have told them in the past 12 months that they had these conditions. Nevertheless, analysts would have the option of deciding how to handle these types of special situations.

7.4.10 Adult Mental Health Service Utilization Module

The module on adult mental health service utilization asked adult respondents about (1) their receipt of specific sources of inpatient or outpatient mental health services in the past 12 months, (2) the length of time that respondents spent in specific inpatient mental health settings or the number of outpatient visits that respondents made to specific types of outpatient mental health providers, (3) payment sources for mental health services, (4) use of prescribed medication for a mental health condition, (5) unmet demand for services (i.e., the respondent felt the need for mental health services but did not receive them), (6) use of alternative sources of treatment (e.g., acupuncture), and (7) how respondents were prompted to get treatment. If the lifetime substance treatment question TX01 indicated that respondents had received treatment for their use of alcohol or other drugs (or if they answered the question as "don't know" or "refused"), respondents were instructed not to include substance abuse treatment in their answers about mental health treatment.

Sources of inpatient mental health treatment or counseling that were asked about in the module included (1) a private or public psychiatric hospital, (2) a psychiatric unit within a general hospital, (3) a medical unit within a general hospital, (4) another type of hospital, (5) a residential treatment center, or (6) "some other type of facility." Sources of outpatient mental health treatment or counseling that were asked about in the module included (1) an outpatient mental health clinic or center, (2) the office of a private therapist not associated with a clinic, (3) a doctor's office that was not part of a clinic, (4) an outpatient medical clinic, (5) a partial day hospital or day treatment program, or (6) "some other place." Sources of alternative treatment that were asked about in the module included treatment from (1) an acupuncturist or acupressurist; (2) a chiropractor; (3) an herbalist; (4) an in-person support group or self-help group; (5) an Internet support group or chat room; (6) a spiritual or religious advisor, such as (but

not limited to) a pastor, priest, or rabbi; (7) a telephone hotline; (8) a massage therapist; or (9) "other" (i.e., some other source).

An important aspect of processing the variables in this section involved assignment of legitimate skip codes, where relevant (Section 2.4.2). That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 12 to 17, and (2) assignment of legitimate skip codes to adult respondents' data based on routing logic within the adult mental health service utilization module. For example, if respondents reported that they did not stay overnight or longer in a hospital or other facility to receive mental health counseling in the past 12 months (AUINPYR = 2), all subsequent variables pertaining to inpatient mental health services were assigned legitimate skip codes.

In addition, if respondents did not report receiving treatment in a particular facility or setting in the past 12 months, the questions pertaining to the number of times they received treatment in that setting were skipped. For example, if respondents reported receiving outpatient mental health services in the past 12 months (AUOPTYR = 1) but did not indicate that they received outpatient services in a day treatment program, the edited variable pertaining to receipt of day treatment services (AUOPDTMT) was assigned a legitimate skip code. If respondents in 2012 reported receiving outpatient services in one or more locations from the lists they were provided but they did not report receiving services in "some other place," the edited "OTHER, Specify" variable AUOPYRSP for outpatient services in some other place was assigned a legitimate skip codes.

Similarly, if respondents reported only one source of payment for inpatient or outpatient mental health services, there was no need to ask them who paid for (or would pay for) most of the inpatient or outpatient services that they received. For example, if respondents reported that they received outpatient mental health services in the past 12 months but reported only that private insurance paid for their outpatient mental health services, the edited variable pertaining to the principal payment source (AUPOPMOS) was assigned a legitimate skip code.

In questions pertaining to the specific places where respondents received inpatient or outpatient mental health services in the past 12 months, they were allowed to enter more than one place from the list where they received services. Similarly, respondents could select more than one response from lists of payment sources for their inpatient or outpatient services. Information for each of these mental health service locations or payment sources was subsequently captured as a discrete variable. For example, information about receipt of inpatient mental health services in a psychiatric hospital, the psychiatric unit of a general hospital, the medical unit of a general hospital, another type of hospital, a residential treatment center, or some other type of facility was captured in the variables AUINPSYH, AUINPGEN, AUINMEDU, AUINAHSP, AUINRESD, and AUINSFAC, respectively. These individual edited variables corresponding to these "enter all that apply" questions in the adult mental health service utilization module were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (see Section 2.4.4).

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report what specific places they received mental health services or what specific sources paid (or would pay) for their mental

health treatment (Section 2.2.2). If an entire list was blank but respondents had previously reported receiving inpatient services (e.g., if respondents broke off the interview), then the lists of variables pertaining to locations for inpatient services or payment for inpatient services retained a code of 98 (i.e., "blank"); similar logic was applied if respondents reported receiving outpatient mental health services but the location or payment variables were entirely blank.

Adult respondents also were asked if there was any time in the past 12 months when they felt the need for mental health treatment but did not get services. Respondents who answered this question as "yes" then were asked to indicate the reason (or reasons) why they did not get treatment. Based on a review of what respondents had specified in quarter 1 of 2002 as leading "other" reasons for not getting mental health treatment, an additional question (ADMT27A) has been included in the module since 2003 to capture commonly endorsed other reasons for not getting treatment. Therefore, if respondents answered question ADMT26 as "yes" (i.e., perceived the need for mental health treatment), they were routed first to question ADMT27, where they could report up to nine reasons for not receiving mental health treatment, including "some other reason or reasons." If respondents chose the "some other reason or reasons" option in question ADMT27, they were routed next to question ADMT27A, where they could choose additional reasons for not getting mental health treatment; again, respondents were given the option in question ADMT27A to report "some other reason or reasons." If respondents chose this "other" response category in ADMT27A, they were asked to specify the *most important* other reason why they did not get treatment. Respondents since 2003 also have been asked to specify only the most important other reason for not getting mental health treatment.

Questions ADMT27 and ADMT27A were "enter all that apply" questions, in which respondents could choose more than one reason from each list. Each response option (e.g., "You couldn't afford the cost") was captured as a separate variable. The edited variables corresponding to the individual response options in ADMT27 and ADMT27A were coded as 1 (Response entered) or 6 (Response not entered), if at least one item was chosen from the lists in ADMT27 or ADMT27A (Section 2.4.4).

In addition, the ninth response option in question ADMT27 ("some other reason or reasons") was considered principally to be a "toggle" to question ADMT27A. Therefore, separate "some other reason or reasons" variables were not created to correspond to the last response category in ADMT27 and the last category in ADMT27A.

The edits described below have been implemented since 2003 to take into account the data from question ADMT27A.

- If respondents chose the ninth response option in ADMT27, any response that was entered from the ADMT27A series was coded as 1, and anything that respondents did not choose from the ADMT27A list was coded as 6.
- If respondents chose a response from ADMT27 but did not choose the ninth response category in ADMT27, the variables corresponding to the response categories in ADMT27A (i.e., AUUNOND through AUUNSOR) all were given a code of 6 (Response not entered), rather than being assigned "legitimate skip" codes. That is, ADMT27 and ADMT27A were considered together to be one big series of reasons.

- If respondents chose category 9 in question ADMT27, chose at least one reason from ADMT27A, but did not choose category 7 in ADMT27A (some other reason or reasons), the edited variable AUUNSOR (some other reason or reasons for not receiving mental health treatment) was assigned a code of 6 (Response not entered). That is, it was inferred in this situation that the list of specific reasons in ADMT27A was adequate for capturing why respondents did not get mental health treatment. For example, if a respondent chose category 9 in ADMT27 and then chose only category 2 in ADMT27A ("You thought you could handle the problem without treatment"), it would be reasonable to infer that this response in ADMT27A was the only other reason why the respondent did not get treatment.
- If respondents chose response category 9 in question ADMT27, it was possible for them to answer ADMT27A as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable AUUNSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reason or reasons" somewhere in the series. Remaining variables corresponding to the ADMT27A series retained codes of 94 ("don't know") or 97 ("refused").
- If respondents answered question ADMT27 as "don't know" or "refused," question ADMT27A was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the ADMT27A list.
- If AUUNSOR had a value of 6 (see above), the "OTHER, Specify" variable AUUNRIMP was assigned a legitimate skip code. If AUUNSOR had a refusal code, that refusal was propagated onto AUUNRIMP.
- If AUUNSOR had a code of 1 when the respondent answered "don't know" or "refused" to the ADMT27A series, the "OTHER, Specify" variable AUUNRIMP retained a code of 98 (blank).

Consistent with the editing procedures in prior years, if AUUNMTYR indicated that there was not a time in the past 12 months when respondents felt the need for mental health treatment but did not receive services (AUUNMTYR = 2), the edited variables corresponding to questions ADMT27 and ADMT27A were assigned legitimate skip codes. Similarly, if AUUNMTYR was refused, that refusal was propagated onto the skipped variables from questions ADMT27 and ADMT27A.

Also, since 2004, respondents have been shown a list of alternative sources of mental health treatment in question ADMT29A (edited variable AUALTYR) and were asked whether they had received treatment, counseling, or support from other sources such as these in the past 12 months. Respondents who answered ADMT29A as "yes" were asked question ADMT29B, pertaining to specific sources of alternative mental health treatment in the past 12 months. ADMT29B was an "enter all that apply" type of question. Therefore, the individual variables AUALACUP (acupuncturist or acupressurist), AUALCHIR (chiropractor), AUALHERB (herbalist), AUALSGRP (in-person support group or self-help group), AUALINET (Internet support group or chat room), AUALRELG (spiritual or religious advisor, such as a pastor, priest, rabbi), AUALHLIN (telephone hotline), AUALMASG (massage therapist), and AUALOTH (other source) pertaining to the individual sources of alternative treatment in ADMT29B were

assigned codes of 1 or 6, as described previously in this section and in Section 2.4.4, when AUALTYR = 1 (i.e., "yes"). When AUALTYR = 2 (i.e., "no"), AUALCHIR through AUALOTH and AUALOTSP (the "OTHER, Specify" variable for other sources of alternative treatment) were assigned legitimate skip codes. When AUALTYR was refused, that refusal was propagated to AUALACUP through AUALOTSP. When AUALOTH was coded as 6 (Response not entered), a legitimate skip code was assigned to AUALOTSP.

Coding of AUALOTSP, regarding other alternative practitioners, was based on information from the National Institutes of Health's (NIH's) National Center for Complementary and Alternative Medicine (NCCAM) (http://nccam.nih.gov/health/whatiscam/). NCCAM groups complementary health approaches into the following categories:

- *Natural Products*: Herbalists fall under this type of therapy, along with use of dietary supplements and probiotics.
- *Mind and Body Practices*: Forms of mind and body practices include the following
  - acupuncture,
  - massage therapy,
  - meditation techniques,
  - movement therapies, such as Pilates and Rolfing Structural Integration,
  - relaxation techniques, such as breathing exercises and guided imagery,
  - spinal manipulation practiced by health care professionals such as chiropractors and osteopaths,
  - tai chi or qi gong practices of movement from traditional Chinese medicine,
  - yoga, and
  - other mind and body practices, such as healing touch and hypnotherapy.
- *Other Complementary Health Approaches*: These include traditional healers, Ayurvedic medicine from India, traditional Chinese medicine, homeopathy, and naturopathy.

In addition, the "OTHER, Specify" variable AUOPYRSP (other source of outpatient mental health treatment or counseling) had previously included a code 11 for support groups, self-help groups, or group counseling. However, AUALSGRP pertained only to support groups or self-help groups that were not commonly part of treatment or counseling from a mental health professional. In contrast, group counseling likely would be administered from a mental health professional. For this reason, respondents who reported group counseling were assigned to a category in AUOPYRSP that has been present since 2003 (43 = Group counseling, self-help not specified). Only those respondents who reported receiving treatment from support groups or self-help groups retained a code of 11 for AUOPYRSP. Therefore, documentation for code 11 in AUOPYRSP has read as "Support group/self-help group" since 2004. Respondents who reported in AUOPYRSP that they had received treatment or counseling from a support group or self-help group were logically inferred in AUALSGRP to have received treatment from this source, if

AUALSGRP had not already been coded as 1; this issue is described further in Table B.25 in Appendix B.

In subsequent analyses of the adult mental health service utilization data, respondents were not classified as having received outpatient mental health treatment if the only "outpatient" location that they reported was a support group or self-help group. In contrast, group counseling was considered a valid other form of outpatient treatment. Therefore, to facilitate analysis of trends in adults' receipt of outpatient mental health treatment in the past 12 months, the variable AUOPYRSP also was revised in 2003 to reclassify respondents into category 43 if they reported group counseling and to retain a code of 11 in AUOPYRSP only for those respondents who reported receiving services from a support group or a self-help group.

Table B.25 discusses additional edits that were relevant to the adult mental health service utilization variables. For example, respondents could report receipt of outpatient mental health services in "some other place" and then specify a location (e.g., a private therapist's office) that they had not already chosen as a place where they received services. In these situations, respondents were logically inferred to have received services at that location (Section 2.4.5). For example, if respondents had not already indicated that they received outpatient mental health treatment in the office of a private therapist, the edited variable AUOPTHER was assigned a code of 3 (Response entered LOGICALLY ASSIGNED).

7.4.11 Social Environment Module

As noted in Section 7.1, the social environment module was administered only to adults. This section included questions about respondents' changes of residence in the past 5 years, involvement in criminal or potentially criminal activities, attitudes about adults trying marijuana once or twice, and religious involvement. A shorter version of this module has been administered since 2005 with the deletion of items related to neighborhood cohesiveness. Consequently, the name of this module has been changed from "social and neighborhood environment" prior to 2005 to "social environment" since 2005.

As was the case in prior years, minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes for respondents who were aged 12 to 17 (Section 2.4.2). Adults were asked all questions within the social environment module. As noted in Section 7.2, responses were replaced with codes for bad data if adult respondents keyed "1" (or multiple-digit responses of "1" if they occurred) to all of the questions that were asked in the module. Although codes of "1" would be plausible for all variables in this module, responses were set to bad data if the pattern of keying "1" wherever possible continued from prior modules.

7.4.12 Parenting Experiences Module

The parenting experiences module was intended to be administered only in dwelling units (DUs) where (1) two people had been selected for an interview, (2) a 12- to 17-year-old had been selected for an interview (regardless of whether the youth completed the interview), and (3) the respondent being interviewed was the parent or legal guardian of the 12- to 17-year-old who also was selected for an interview. Editing of the parenting experiences data first involved editing the

field interviewer (FI) checkpoint variables (FIPE1, FIPE2, and FIPE3) completed by the interviewers toward the beginning of the interview.[117] The variables in the parenting experiences module then were edited based on the final values assigned to the edited FIPE variables.

The content of this module did not change in 2012. The ACASI "lockout" feature that was described in Section 7.3 did not affect how the FI checkpoint data were edited but did affect the editing of the parenting experiences data. Issues associated with this ACASI "lockout" are described in Section 7.4.12.2.

7.4.12.1 Editing of the Field Interviewer Checkpoint Variables

Interviewers were instructed to enter into these checkpoints the relevant information described above for determining whether respondents were eligible to be administered the parenting experiences questions. These checkpoint variables were edited for consistency with the pair-selection and pair-respondent sample variables (PAIRSEL and PAIRRESP, respectively). These checkpoints were interviewer-administered and not self-administered. Editing of these checkpoints was related to the edits for the parenting experiences questions (which were self-administered), however, because the final values in the edited checkpoints were critical for determining whether respondents were in fact eligible to be asked the parenting experiences questions.

Editing of the FIPE1 Checkpoint (and Related Edits). First, the FIPE1 variable was edited for consistency with the pair-selection variable PAIRSEL. Specifically, this checkpoint pertained to whether two people were selected for an interview at that DU. There were no situations in 2012 when two people were interviewed at a given DU without two people having first been selected. Therefore, editing FIPE1 involved reviewing only information on the number of people selected for an interview at that DU based on PAIRSEL.

If the pair-selection data indicated that two people were selected from that DU, then FIPE1 should have been answered as "yes." Therefore, if the pair-selection data indicated that two people were selected and FIPE1 was not answered as "yes," a code of 3 (i.e., Yes LOGICALLY ASSIGNED) was assigned to the edited FIPE1 variable (SKPX2PER). Similarly, if the pair-selection data indicated that only one person was selected from that DU, then FIPE1 should have been answered as "no." Therefore, if the pair-selection data indicated that only one person was selected and FIPE1 was not answered as "no," the editing procedures logically inferred that "no" should have been the answer. If the edited version of FIPE1 indicated that two people were not selected for an interview, then the edited versions of FIPE2 (SKPX1217) and FIPE3 (SKPXPRNT) were assigned legitimate skip codes. If data existed in FIPE2 or FIPE3 when the edited SKPX2PER was inferred to be answered as "no," SKPX1217 and SKPXPRNT were assigned codes of 89 (i.e., LEGITIMATE SKIP Logically assigned) to signify that these two checkpoints should have been skipped (Section 2.4.2).

Editing of the FIPE2 Checkpoint (and Related Edits). Next, FIPE2 was edited for consistency with PAIRSEL, PAIRRESP, and the age of the respondent. Specifically, this

[117] "Checkpoint" refers to an item completed by the interviewer about the location of the sampled dwelling unit (SDU) or characteristics of the sample within the SDU; these checkpoints are not seen by the respondent and are used to determine what the respondent is asked in subsequent sections of the interview.

checkpoint pertained to whether a 12 to 17 year old was selected for an interview at that DU, *regardless of whether the selected youth actually responded*. Edits of the FIPE2 checkpoint data involved review of both the pair-selection data (PAIRSEL) and the pair-respondent data (PAIRRESP) in case either indicated that a 12 to 17 year old was selected or interviewed.

The age of the respondent was taken into account because interviewers were skipped past this checkpoint if respondents were aged 12 to 17. Therefore, the edited version of FIPE2 (SKPX1217) was assigned legitimate skip codes (i.e., 99 if FIPE2 was blank and 89 if FIPE2 was not blank) when the respondent was a youth.

The remaining edits for FIPE2 were implemented when the respondent was an adult. If both PAIRSEL and PAIRRESP indicated that a 12 to 17 year old was neither selected nor interviewed, it reasonably could be inferred that FIPE2 should have been answered as "no." If FIPE2 was not already answered as "no," the edits assigned a code to SKPX1217 to indicate that a response of "no" was logically inferred. This included situations in which the pair-selection data indicated that a 12 to 17 year old was not selected, and a completed interview was obtained from only one respondent, who was not aged 12 to 17, regardless of whether PAIRSEL and PAIRRESP were totally consistent. For example, if the pair-selection data indicated that an 18 to 25 year old and a 26 to 34 year old were selected, but a single interview was obtained from a 35 to 49 year old, the pair-selection and pair-respondent data were not totally consistent, but neither would suggest that a 12 to 17 year old should have been selected. When the edited SKPX1217 indicated that a 12 to 17 year old was not selected, including situations described above in which the edits inferred that no 12 to 17 year old was selected, then legitimate skip codes were assigned to the edited variable SKPXPRNT corresponding to FIPE3 (code of 99 if FIPE3 was blank; or 89 if it was not blank).

If either PAIRSEL or PAIRRESP indicated that a 12 to 17 year old was selected or interviewed, it could be inferred that FIPE2 should have been answered as "yes." Therefore, if FIPE2 was not already answered as "yes," a special code was assigned to SKPX1217 to indicate that a response of "yes" was logically inferred. This included the following situations: (1) PAIRSEL indicated that a 12 to 17 year old was selected, and PAIRRESP indicated that an interview was obtained from a 12 to 17 year old, regardless of whether PAIRSEL and PAIRRESP matched exactly (e.g., a 12 to 17 year old and a 26 to 34 year old were selected, but interviews were obtained from a 12 to 17 year old and a 35 to 49 year old); and (2) PAIRSEL indicated that a 12 to 17 year old was selected, but a single interview from an adult was obtained at the DU, regardless of whether the adult category from PAIRSEL matched the category in PAIRRESP (e.g., a 12 to 17 year old and 26 to 34 year old were selected, but a single interview was obtained from a 35 to 49 year old). In the latter situation, the respondent result (from PAIRRESP) was not totally consistent with what would be expected based on the pair selection, but PAIRRESP would not provide any information to directly contradict the indication from PAIRSEL that a 12 to 17 year old was selected.

If PAIRSEL and PAIRRESP disagreed when two people were interviewed, with one indicating the selection or interview of a 12 to 17 year old but the other variable did not, then special codes were assigned to SKPX1217. When this type of inconsistency occurred, a code of 11 was assigned to SKPX1217 when FIPE2 was originally answered as "yes," and a code of 12 was assigned when FIPE2 was originally answered as "no."

Suppose, for example, that PAIRSEL indicated that a 12 to 17 year old and a 35 to 49 year old were selected for the interview, but PAIRRESP indicated that an 18 to 25 year old and a 35 to 49 year old were actually interviewed, with the interviewer keying FIPE2 = 1 in the adult's interview (i.e., "yes," a 12 to 17 year old was selected for an interview at this DU). In this situation, the "yes" in FIPE2 was consistent with who was *selected* (according to the information provided by the screening respondent), but it was not consistent with the ages provided by the respondents themselves. Therefore, the edited variable SKPX1217 would be set to a value of 11 in this example.

This latter edit preserved the information that the interviewer originally entered but also denoted that an inconsistency existed between PAIRSEL and PAIRRESP. This edit also was designed to preserve any possible parenting experiences data when both FIPE2 and FIPE3 (see below) were answered as "yes" but there was an inconsistency between PAIRSEL and PAIRRESP. When an inconsistency occurred between PAIRSEL and PAIRRESP, an analyst would have discretion about whether to use parenting experiences data in an analysis.

Editing of the FIPE3 Checkpoint. This checkpoint pertained to whether the respondent was the parent or legal guardian of the 12 to 17 year old who also was selected to be interviewed at that DU. Respondents who were aged 12 to 17 were skipped out of FIPE3 and did not have an opportunity to be routed into the parenting experiences module. Therefore, when FIPE3 had been skipped because the respondent was 12 to 17, the edited FIPE3 variable SKPXPRNT was assigned a legitimate skip code.

No further editing of FIPE3 was done when PAIRSEL indicated that a 12 to 17 year old was selected and PAIRRESP had some result *other than* that of two adults having been interviewed at that DU. The rationale for this approach was that FIPE3 was based on who the actual respondent was, provided that a 12 to 17 year old was selected. For example, if PAIRSEL indicated that a 12 to 17 year old and a 26 to 34 year old were selected, but a 35 to 49 year old and a 12 to 17 year old were interviewed, and FIPE3 was answered as "yes" (i.e., this adult respondent is the parent of the youth who was selected), that 35- to 49-year-old respondent may indeed have been a parent or legal guardian of the youth who was selected. This principle also would have held if the selected youth did not respond. Therefore, any data that were present in the parenting experiences module would be preserved.

In contrast, the following situations could occur when FIPE3 was inconsistent with either PAIRSEL or PAIRRESP: (1) PAIRSEL indicates that a youth/adult pair was selected, but two adult interviews were obtained at that DU; or (2) PAIRRESP indicated that a youth/adult pair was interviewed, but PAIRSEL indicated that an adult/adult pair was selected. When either of these inconsistencies occurred, a code of 11 was assigned to SKPXPRNT when FIPE3 was originally answered as "yes," and a code of 12 was assigned when FIPE3 was originally answered as "no."

Suppose, for example, that PAIRSEL indicated that an 18 to 25 year old and a 35 to 49 year old were selected for the interview, but PAIRRESP indicated that a 12 to 17 year old and a 35 to 49 year old were actually interviewed, and the interviewer keyed FIPE2 = 1 and FIPE3 = 1 in the adult's interview. Stated another way, the interviewer indicated that "yes," a 12 to 17 year old was selected for an interview at this DU, and "yes," this 35- to 49-year-old respondent was

the parent of the 12- to 17-year-old youth who was selected. In this situation, FIPE3 was consistent with PAIRRESP but not PAIRSEL. Furthermore, based on who was interviewed at that DU, the 35 to 49 year old may indeed be the parent of the 12 to 17 year old who also was interviewed at that DU. In this situation, the edited SKPXPRNT would be set to a value of 11 to denote that this type of inconsistency has occurred. Again, this edit would preserve any possible parenting experiences data—especially in situations in which an adult/child *respondent* pair was obtained.

7.4.12.2 Editing of the Variables in the Parenting Experiences Module

The variables in the actual parenting experiences module were edited according to the final values assigned to SKPX2PER, SKPX1217, and SKPXPRNT based on the edits described in Section 7.4.12.1. In particular, if these three variables indicated that the respondent was not eligible to be administered the parenting experiences questions, then the edits assigned the appropriate legitimate skip codes to the parenting experiences variables. This included replacing blank values with legitimate skip codes when a code of 12 had been assigned SKPXPRNT and the parenting experiences module has been skipped. The rationale for this latter edit was that even if FIPE3 was answered as "no" when PAIRSEL and PAIRRESP were inconsistent, the adult respondent still may not have been the parent or legal guardian of the youth who also was selected for an interview at that DU.

Conversely, if a respondent had been skipped out of the parenting experiences module and the edited FIPE variables SKPX2PER, SKPX1217, or SKPXPRNT indicated that the respondent was *potentially* eligible to be administered the parenting experiences questions (i.e., the respondent skipped the module based on the original answers in the FIPE questions, but other data suggested that the respondent may have been eligible to be asked these questions), then the edited parenting experiences variables retained a value of "blank." For example, if FIPE2 had been keyed as "no" and it was inferred for SKPX1217 that a 12 to 17 year old was selected (i.e., SKPX1217 = 3), then FIPE3 and the parenting experiences questions also would have been skipped. In this situation, the respondent's eligibility or ineligibility to be administered the parenting experiences questions could not be determined because the field interviewer (FI) was not routed to the final checkpoint. Therefore, it could not be determined whether the respondent should have been asked the parenting experiences questions or should have been skipped.

As noted previously, the interview has included a "lockout" feature since 2003 that does not allow interviewers or respondents to go back into the ACASI sections and change their answers once that section of the interview had been completed. However, interviewers could go back to the beginning of the interview after respondents had been administered the ACASI sections and change FIPE1 through FIPE3 in a manner that made the final value in FIPE3 inconsistent with the presence of data in the parenting experiences module. Specifically, FIPE3 ("Is this respondent the parent or legal guardian of the 12 - 17 year old child who was selected for an interview?") could be answered as something other than "yes," with at least some data existing in the parenting experiences module. According to the CAI logic, however, the module was to be administered only when the interviewer indicated that the respondent was the parent or guardian of the selected 12 to 17 year old (i.e., FIPE3 = 1).

The following is an example of a scenario where parenting experiences data could exist when FIPE3 was not answered as "yes." If the FI initially answered FIPE1 through FIPE3 as "yes" (i.e., two people were selected at that DU, a 12 to 17 year old was selected at that DU, and the respondent is the parent/guardian of that 12 to 17 year old), the respondent would be routed through the ACASI parenting experiences questions. In the household roster section of the "back-end" demographics section, however, a "hard error" would be triggered if the second person selected to be interviewed was not identified in the roster. Before the interview could proceed, the interviewer would need to change the information in the household roster to make it consistent with the information in FIPE1 through FIPE3, or else the interviewer would need to go back and change the information in FIPE1 through FIPE3 to make it consistent with the roster. In particular, an interviewer could resolve this inconsistency by going back and changing one of the answers in FIPE1 through FIPE3 from "yes" to "no." Because the interviewer and respondent are locked out of ACASI, however, the parenting experiences data would be saved as it had been entered originally. The instrument development team was able to reproduce this type of scenario to yield a result where FIPE3 was not answered as "yes" but data existed in one or more questions in the parenting experiences module.

If interviewers changed the value in FIPE3 to "no," this would not present a problem in editing the parenting experiences data because the corresponding edited variable SKPXPRNT indicated that the respondent was not the parent or legal guardian of the youth who was selected for an interview. Consequently, parenting experiences data were edited to infer that these respondents should have legitimately skipped the parenting experiences module. Any data that existed in the parenting experiences module were overwritten with codes of 89 (or 9989, etc.).

However, changes made by the interviewer to FIPE1 through FIPE3 that resulted in SKPXPRNT having a final value of 98 (i.e., blank) were more problematic. Because the edits for SKPX2PER, SKPX1217, and SKPXPRNT were consistent with the pair-selection and pair-respondent data (PAIRSEL and PAIRRESP, respectively), the decision was made to retain the value of 98 in SKPXPRNT. Therefore, any nonblank values that existed in the parenting experiences module were replaced with codes for bad data.

One set of variables involved skip logic within the parenting experiences module. Specifically, respondents were skipped out of question PE04 (length of most serious discussion about the dangers of tobacco/alcohol/other drug use) when question pe03 had a value of 1 (i.e., talked with child 0 times in the past year about the dangers of tobacco/alcohol/other drug use), or if PE03 was answered as "don't know" or "refused." Standard procedures for assigning legitimate skip codes (Section 2.4.2) or propagating refusal codes (Section 2.4.3) were implemented in the edited version of question PE04 (PXSERDIS) depending on the response in PE03 (edited variable PXKIDYR).

Parents were asked to report the birth date of the youth who was selected for an interview at that DU (question PE01). However, the birth year that respondents could enter for the youth in question PE01 was restricted to ages that would be more consistent with selection of a 12 to 17 year old (but also allowed for birth dates that would include 18 year olds, in case a 17-year-old respondent just recently had a birthday). Thus, respondents were prevented from entering birth dates that would be extremely inconsistent with selection of a 12 to 17 year old (such as entry of the current interview year for the birth year).

A refinement has been implemented for the parenting experiences edit logic since 2003 to take into account the situation in which all remaining parenting experiences questions had been skipped because the respondent did not provide a date of birth for the selected youth in question PE01 or did not provide an age for the youth in question PE01B. Prior to 2003, the skipped parenting experiences variables were assigned legitimate skip codes. Since 2003, the parenting experiences variables have retained codes for "blank" when this pattern occurred. The effect of this refinement in 2003 was to make the frequencies of legitimate skip values in PXCHCIG (corresponding to question PE02) and subsequent parenting experiences variables agree with the total count of codes of 2 or 12 (i.e., "no"), or 89 or 99 (i.e., legitimate skip) in SKPXPRNT.

The CAI program also calculated an age for the youth who was selected for an interview based on the youth's date of birth (as reported by the parent) and the interview date at the start of the parenting experiences module. Respondents were asked to confirm this age (question PE01A). If parents did not confirm the age that the CAI program calculated for the youth, they were asked to provide a corrected age for the youth who was selected for an interview (question PE01B). Similarly, if respondents did not know or refused to report the date of birth of the selected youth, they were asked to report an age in question PE01B without having to indicate the youth's date of birth.

This information was captured in the created variable PXCHLDAG. Specifically, PXCHLDAG contained the age based on the reported date of birth for the youth and the interview date (if respondents confirmed that this age was correct), or else PXCHLDAG contained the age supplied by the respondent from question PE01B. If respondents supplied a corrected age for the youth in question PE01B that was between 12 and 18 and it mismatched the age of the youth that was calculated from the birth date and interview date information, the edited variables containing the birth date information for the youth (PXBMONTH, PXBDAY, and PXBYR) were assigned bad data values. If respondents answered question pe01b as "don't know" or "refused" when they were asked to provide a corrected age for the selected youth, that response of "don't know" or "refused" was assigned to PXCHLDAG. In addition, if respondents answered question PE01A as "don't know" or "refused" when they were asked to confirm the age of the youth who was selected for the interview, the age for the youth that the CAI program had previously calculated was retained in PXCHLDAG (see above). When values in parenting experiences variables had been set to bad data because SKPXPRNT was blank, PXCHLDAG also was assigned a code for bad data.

A recoded variable (PXCMPAGE, for "compare age") also was created that compared the selected youth's age (from PXCHLDAG) with the respondent's age for the second interview conducted at that DU. If two interviews were obtained at that DU and a 12 to 17 year old was selected for an interview, then PXCMPAGE was calculated as the absolute value of the difference between PXCHLDAG and the actual age of the second respondent, within defined categories (i.e., 0 year difference in ages; 1 year difference in ages; 2 year difference in ages; 3 to 4 year difference in ages; and 5 or more year difference). If the adult respondent answered "don't know" or "refused" to the question about the youth's date of birth, or if the youth's date of birth information was set to bad data because of invalid dates, these codes were reflected in PXCMPAGE.

For the large majority of cases where an interview was obtained from a 12 to 17 year old, PXCMPAGE indicated no difference between the age based on the date of birth reported by the parent and the youth's age recorded in the second interview at that DU. Nevertheless, information about more extreme differences in ages as recorded by PXCMPAGE (e.g., a difference of 2 or more years between the two ages) could be used by analysts in deciding whether to use the parenting experiences data in an analysis. When the second interview was from an 18 year old, PXCMPAGE was assigned a value of 18. When the second interview was from an adult older than age 18 (i.e., and the parent was supposed to be reporting about a 12 to 17 year old), the edit program assigned a code of 50 to PXCMPAGE. Again, these codes were designed to give analysts discretion in using or disregarding parenting experiences data when the second interview at a DU came from an adult.

If a 12 to 17 year old was supposed to be selected at a given DU but only the adult was interviewed, PXCMPAGE was assigned a code of 93. This code was assigned because there were no data to corroborate the youth's date of birth reported by the parent.

If the edited FIPE variables from above indicated that the respondent was not eligible to be administered the parenting experiences questions, then PXCMPAGE was assigned a code of 99 (i.e., legitimate skip). That included situations in which the edited FIPE3 was assigned a code of 12 because of an inconsistency between PAIRSEL and PAIRRESP, and the parenting experiences module had been skipped (see above). Otherwise, if the parenting experiences module was all blank or if PXCMPAGE was undefined for some other reason, then PXCMPAGE was assigned a code of 98. This code of 98 in PXCMPAGE meant "other missing." This code of 98 also was applied in PXCMPAGE when SKPXPRNT was blank and parenting experiences data had been replaced with codes for bad data.

7.4.13 Youth Experiences Module

As noted in Section 7.1, the youth experiences module was administered only to respondents aged 12 to 17. This section included questions about changes of residence in the past 5 years; school enrollment and related issues (e.g., opinions about the importance of assigned schoolwork) in the past 12 months, including homeschooling; other social and family characteristics (e.g., substance use behaviors of other students or friends, personal attitudes about substance use, parental attitudes about substance use); people with whom the youth could confide about a serious problem; exposure to alcohol- and other drug-related prevention messages in school or outside school; and personal behaviors (e.g., involvement in criminal or potentially criminal activities, involvement in extracurricular activities) that might be associated positively or negatively with the use of alcohol or other drugs. The youth experiences module also included questions about youths' religious involvement in the past 12 months and opinions about religious issues.

Minimal processing of data was done to variables in this section. The primary data processing involved assignment of legitimate skip codes based on the CAI routing logic. That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 18 or older, and (2) assignment of legitimate skip codes to youths' data based on routing logic within the youth experiences module.

Some special issues were encountered in editing the variables corresponding to question YE22, which pertained to people whom youths could turn to if they had a serious problem. Specifically, youths were asked to enter all the different types of people to whom they could turn to (e.g., a parent, a friend). This question also included a response category for youths who felt that there was no one they could talk to about a serious problem.

The questions indicating the youths' relationships to people whom they could turn to if the youths had a serious problem were "enter all that apply" questions. The individual edited variables for relationships were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (see Section 2.4.4). If the entire list of responses was blank (e.g., if a youth broke off the interview before getting to these questions), the edited variables retained a code of "blank."

Youths could indicate that there was no one they could talk to about a serious problem but then indicate that they could talk to one or more of the people or types of people in the list from question YE22. In this situation, the variable pertaining to the first item in the list ("There is nobody I can talk to about a serious problem") was assigned a code of 11 (if that response was chosen along with another response from the list). Similarly, codes of 11 were assigned to the edited relationship variables (e.g., my mom, my dad) when they were chosen along with the response that there was nobody that the youth could talk to.

7.4.14 Mental Health Module

From 2004 to 2007, the mental health module for adults consisted of two primary components. First, a 12-month Kessler-6 (K6) distress scale was administered, and then questions about lifetime and 12-month major depressive episode (MDE) were asked. Since 2008, however, the K6 questions have collected data on distress in the past 30 days and in the past 12 months (see Section B.4.5 in the 2008 national findings report for details) (Office of Applied Studies, 2009). Respondents were routed into the 12-month version of the K6 if they reported having a period in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt in the past 30 days.

In 2008, adult respondents also were administered one of two impairment scales—an abbreviated World Health Organization Disability Assessment Schedule (WHODAS) (see Section B.4.6 in the 2008 national findings report for details) or the Sheehan Disability Scale (SDS)—and suicidal ideation questions. These impairment and suicide questions were placed after the K6 questions, but before the MDE questions. A random split-sample design was implemented in 2008 for adults where respondents in sample A were administered the WHODAS scale and respondents in sample B were administered the SDS. All adult respondents were administered the suicidal ideation questions after the impairment items, but before the MDE items. Starting with 2009, the SDS items were no longer included in the mental health module, and the WHODAS has been used to assess impairment.

Apart from propagating refusals from the lead questions for the WHODAS, minimal processing of data was done to the mental health variables. The primary data processing involved assignment of legitimate skip codes for respondents who were aged 12 to 17. Legitimate skip codes also were assigned to data from adults based on other skip logic within the module. If

adults broke off the interview before they reached the WHODAS questions, then all of the WHODAS variables retained a code of 98 (i.e., "blank").

7.4.15 Adult and Adolescent Depression Modules

The adult and adolescent depression modules have been present in the interview since 2004. Questions in these modules were based on those used in the National Comorbidity Survey Replication for adults (NCS-R) and the National Comorbidity Survey Replication Adolescent Supplement (NCS-A) (http://www.hcp.med.harvard.edu/ncs/). These depression modules were included in NSDUH to produce lifetime and 12-month prevalence estimates of major depressive episode (MDE), severity of 12-month MDE, age at first MDE, lifetime number of episodes, current and 12-month treatment, and the respondent's perception of treatment effectiveness.

There were some differences in wording between related items in these modules, such as use of simpler wordings for the adolescent depression questions. For example, question AD17 in the adult depression module asked respondents how "severe" their "emotional distress" was during their worst periods lasting 2 weeks or longer when they had problems with their mood. The corresponding question YD17 from the adolescent depression module asked adolescents how "strong" their "bad feelings" were during these periods. Despite these differences in wording, similar naming conventions were used for the variables in these two modules. For example, the edited variable corresponding to question AD17 from the adult depression module was ADWRDST (where WR = worst period, and DST = distress), and the edited variable corresponding to question YD17 from the adolescent depression module was YOWRDST, even though YD17 did not ask about distress. Thus, the only difference in the names for analogous edited variables in these modules was in the use of the two-letter prefix that defined which module a given variable came from: "AD" for variables from the adult depression module and "YO" for adolescent depression variables.

There also were differences in how the CAI program created indicators of MDE for adolescents and adults based on differences in the NCS-A and NCS-R. Specifically, the criteria for defining respondents as having the symptom of loss of interest or pleasure in most things was less restrictive for adolescents than for adults. In particular, the DSM-IV criteria for MDE (American Psychiatric Association, 1994) place more emphasis for adolescents on the cognitive aspects of depression, such as boredom or apathy, rather than on somatic or physical complaints, such as sleep loss, that may be manifest in adults with MDE. For example, somatic or physical complaints, such as sleep loss, may be due to factors during adolescence other than depression. Consequently, somatic or physical complaints that may be associated with MDE among adults function less well as indicators of MDE among adolescents than do cognitive indicators. For this reason, the CAI logic gave adolescents additional opportunities to be classified as having the symptom of loss of interest based on their answers to questions that were not taken into account in classifying adults as having this symptom.

Despite these differences between the adult and adolescent depression modules, the basic logic for asking questions was similar between the two modules. Therefore, the remainder of this section discusses edits for both of these modules together. Except where differences are discussed in terms of how variables were edited for these modules, the same basic edits discussed below applied to variables in both modules.

An important aspect of the processing of variables in these modules consisted of assigning legitimate codes based on the routing logic within these modules. In particular, adults were assigned legitimate skip codes to the edited variables in the adolescent depression module, and adolescents were assigned legitimate skip codes to the edited variables in the adult depression module.

As an additional example, respondents were asked a series of questions to identify changes in appetite or weight. They first were asked whether they had a much smaller appetite than usual during the most recent time when their problems were the worst (questions AD26A and YD26A, corresponding to edited variables ADWRELES and YOWRELES, respectively). If respondents answered the relevant question as "yes," they were skipped out of subsequent questions about increases in appetite and weight gain. If respondents reported having less appetite, they were asked whether they lost weight without trying to, and if so, whether their weight loss was due to them being sick or on a diet; respondents who indicated that their weight loss was due to sickness or a diet were not asked to report how many pounds they lost in weight.

Conversely, if respondents did not report in AD26A/YD26A that they had less appetite than usual, they were asked whether they had a much larger appetite than usual (questions AD26B and YD26B, corresponding to edited variables ADWREMOR and YOWREMOR, respectively). Respondents who reported that they had a much larger appetite were skipped out of remaining questions related to weight loss. These respondents subsequently were asked whether they gained weight without trying to. If respondents reported gaining weight without trying to, the CAI program asked follow-up questions to rule out weight gains due to growth (for respondents aged 21 or younger) or pregnancy (for females); respondents who indicated that they gained weight because they were growing or because they were pregnant were not asked to report how many pounds they gained. Thus, editing of the adult and adolescent depression variables related to changes in appetite or weight involved assignment of legitimate skip codes to these variables based on the routing logic for the corresponding questions.

As noted in Section 2.4.3, legitimate skip codes generally were not assigned if a lead question was answered as "don't know" or "refused." However, important exceptions to this principle were made in editing of the adult and adolescent depression variables because of consideration of other aspects of the routing logic in these modules. In particular, the lead screening questions ASC21 through ASC23 (corresponding to edited variables ADDPREV, ADDSCEV, and ADLOSEV) and YDS21 through YDS23 (corresponding to edited variables YODPREV, YODSCEV, and YOLOSEV) at the beginning of these modules had a special skip logic. If a particular lead question was not answered affirmatively, this logic routed respondents into follow-up questions that could screen respondents into further questions about depression.

In the adult depression module, for example, if question ASC21[118] was answered as "don't know" or "refused," the subsequent question AD01[119] was skipped. In this situation,

[118] Question ASC21 asked, "Have you ever in your life had a period of time lasting several days or longer when most of the day you felt sad, empty, or depressed?"

[119] If respondents answered question ASC21 as "yes," they were asked question AD01. Question AD01 asked, "During times when you felt sad, empty, or depressed most of the day, did you ever feel discouraged about how things were going in your life?"

however, respondents were routed to follow-up question ASC22.[120] Thus, if respondents answered question ASC21 "don't know" or "refused" but they answered question ASC22 as "yes," they were still eligible to be administered the remainder of the adult depression module, depending on how they answered subsequent questions.

Therefore, legitimate skip codes were assigned (where relevant) to the variables corresponding to questions AD01 through AD09 in the adult depression module (edited variables ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK) if at least one item from questions ASC21, ASC22, or ASC23 was answered as "yes" or "no." For example, if question ASC21 was answered as "don't know," and question ASC22 or ASC23 was answered as "yes" or "no," the editing procedures assigned a legitimate skip code to ADDPDISC, corresponding to question AD01. Similarly, if question ASC22 was answered as "don't know" but ASC23 was answered as "yes" or "no," the editing procedures assigned a legitimate skip code to ADDSLSIN, corresponding to question AD02. In turn, if ADDPDISC, ADDPLSIN, and ADDSLSIN all were answered as "no" or had legitimate skip codes after the above edits, then ADLSI2WK (corresponding to AD09) was assigned a legitimate skip code. The values in ADDPDISC, ADDPLSIN, ADDSLSIN, and ADLSI2WK determined whether subsequent variables were assigned legitimate skip codes. The logic provided in this example for adult depression also was applied to the variables in the adolescent depression module.

As discussed in Section 2.4.3, if a lead question that governs a skip pattern was refused, the editing procedures typically "propagated" that refusal from the lead question to the variables that had been skipped. For most of the adult and adolescent depression variables, however, this refusal propagation was not performed. The CAI program contained routines for scoring the symptom indicators for MDE. The CAI program coded the symptom score variable DSMMDEA2 as 2 if the sum of the numbers of codes of 1 (i.e., has symptom), "don't know," or "refused" in the individual symptom indicators was less than 5. Therefore, not propagating refusals helped to avoid situations in which a different overall score might be obtained if analysts were to calculate DSMMDEA2 based on edited variables.

The exception to this rule of not propagating refusals in the adult and adolescent depression modules concerns the final questions regarding receipt of counseling from a medical doctor or other professional about the respondents' symptoms of depression (e.g., questions AD86 through AD86F in the adult depression module). For example, if question AD86 (edited variable ADSEEDOC) was refused, the editing procedures still propagated that refusal code to the skipped variables that were dependent on AD86.

In addition, the CAI program created MDE symptom variables and overall MDE symptom scores for adults and adolescents. Table 7.3 lists the final, edited variables that were created from these symptom variables and overall symptom scores. For each variable, explanation of the meaning of that variable also is provided. The only editing that was done to these variables in Table 7.3 was to assign legitimate skip codes based on the respondent's age (i.e., 12 to 17 or 18 or older). Thus, the values that were created by the CAI program were preserved in the variables listed in Table 7.3.

[120] Question ASC22 asked, "Have you ever had a period of time lasting several days or longer when most of the day you were very discouraged about how things were going in your life?"

Table 7.3 Depression Symptom and Score Variables

Adult Depression Variable	Adolescent Depression Variable	Explanation
AD_MDEA1	YO_MDEA1	Respondent (R) felt sad, empty, depressed, or discouraged most of the day.
AD_MDEA2	YO_MDEA2	R lost interest or pleasure in most things.
AD_MDEA3	YO_MDEA3	R had changes in appetite or weight (not due to growth, pregnancy, illness, or dieting).
AD_MDEA4	YO_MDEA4	R had sleep problems.
AD_MDEA5	YO_MDEA5	Others noticed that the R was restless or lethargic.
AD_MDEA6	YO_MDEA6	R felt tired or low on energy nearly every day.
AD_MDEA7	YO_MDEA7	R felt worthless nearly every day.
AD_MDEA8	YO_MDEA8	R was unable to concentrate or make decisions.
AD_MDEA9	YO_MDEA9	R was suicidal (had thoughts of suicide, made plans, or made an attempt).
ADSMMDEA	YODSMMDE	Score of symptom indicators 1 through 9 from above.

Relatively little additional editing was done to the adult and adolescent depression variables, aside from assigning legitimate skip codes. Additional editing issues that were relevant to these modules are described in the remainder of this section.

If respondents reported a period of time when their symptoms or problems were the worst, they were asked to report how old they were when this time started (edited variables ADWRAGE for adults and YOWRAGE for adolescents). In addition, if respondents scored as positive for MDE (edited variables ADSMMDEA and YODSMMDE) and they reported that these problems caused some interference with their work, social life, or relationships, they were asked to report the age at which these problems first occurred (edited variables ADPBAGE and YOPBAGE). If respondents reported an age of onset in any of these variables that was greater than their current age, these variables were set to bad data. For adults, if ADWRAGE had been set to bad data and the respondent's original answer was age 22 or greater, the respondent was skipped out of the question about weight gain because the respondent was growing (edited variable ADWRGROW). Therefore, if ADWRAGE had been set to bad data, ADWRGROW retained a code of blank.

As noted previously, if respondents reported gaining or losing weight and these gains or losses could not be attributed to factors other than depression (e.g., growth, pregnancy, dieting), respondents were asked to report the number of pounds they gained or lost. In particular, respondents were allowed to report that they gained or lost 0 pounds. No editing was done to the variables ADWRGNLB, ADWRLSLB, YOWRGNLB, or YOWRLSLB when this response of 0 pounds occurred because respondents did not have an opportunity to report gains or losses of less than 1 pound. Furthermore, only gains or losses of 10 or more pounds resulted in respondents being scored as having the symptom of changes in appetite or weight.

In addition, a feature of the logic for asking respondents about gains or losses in weight was that if respondents were asked the question about gaining weight without trying and they answered it as "don't know" or "refused," they had an additional opportunity to be asked questions about losing weight without trying. That is, the program was looking for the first affirmative set of answers that would allow a determination to be made of whether respondents

gained or lost enough pounds to qualify for being depressed. Consequently, no editing was done if respondents originally gave an answer in the questions corresponding to ADWREMOR or YOWREMOR (i.e., having a much larger appetite than usual), they answered the weight gain question (corresponding to ADWRGAIN or YOWRGAIN) as "don't know" or "refused," and then they were routed into the questions about weight loss.

If respondents reported that they talked to a medical doctor or other professional in the past 12 months about the problems they were experiencing related to depression, they were asked to report which professionals they saw or talked to. In the question pertaining to the specific professionals that they saw or talked to, respondents were allowed to enter more than one type of professional from the list they were presented. As in other modules, the "enter all that apply" variables in the adult and adolescent depression modules were coded as 1 if the response was entered and as 6 if the response was not entered (see Section 2.4.4).

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report the specific professionals that they saw or talked to about their problems. If the entire list was blank but respondents had previously reported that they saw or talked to a professional about their problems, then the specific variables corresponding to categories of helping professionals retained a code of 98 (i.e., "blank").

Respondents could report that they saw or talked to "another type of helping professional" and then specify a helping professional that they had already been asked about, such as a psychiatrist. Thus, for example, if the edited variable ADPSYMD, pertaining to services from a psychiatrist, was not coded as 1 and respondents specified that they saw or talked to a psychiatrist, ADPSYMD was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED.

7.4.16 Youth Mental Health Service Utilization Module

The module on youth mental health service utilization asked respondents aged 12 to 17 about their receipt of specific sources of inpatient, foster care, outpatient, or school-based mental health services in the past 12 months; the number of nights that respondents spent in specific inpatient or foster care mental health settings; the number of times they visited specific types of outpatient or school-based mental health providers; and the reasons for receiving inpatient, foster care, outpatient, or school-based services for mental health problems the last time they received such services. Specific sources of mental health services that respondents were asked about included (1) any type of hospital, (2) a residential treatment center, (3) foster care or a therapeutic foster home, (4) a partial day hospital or day treatment program, (5) a mental health clinic or center, (6) a private therapist, (7) an in-home therapist, (8) a pediatrician or other family doctor, (9) special education services, and (10) in-school counseling, such as from school counselors or school psychologists.

In 2009, a set of questions from prior years was deleted from the youth mental health utilization module; the deleted questions asked about receipt of special education services and receipt of school counseling in the past 12 months, and whether youths were ever in jail or foster care. In their place, a set of questions has been included since 2009 that measure the same

constructs as the questions from prior years but with different wordings. For youths' experiences with the juvenile justice system, the questions since 2009 also focus on the past 12 months rather than the lifetime period and collect additional information about these encounters with the juvenile justice system. The revised questions that have been included since 2009 asked about the following for the past 12 months: (1) receipt of counseling or treatment for mental health problems from social workers, school psychologists, or school counselors; (2) the reason for the last talk with a social worker, school psychologist, or school counselor (if applicable); (3) attendance at a school for students with emotional or behavioral problems or participation in a special program at a regular school for students with emotional or behavioral problems; (4) time spent overnight or longer in a juvenile detention center; (5) the number of nights spent in a juvenile detention center (if applicable); and (6) receipt of treatment or counseling for emotional or behavioral problems in a juvenile detention center (if applicable).

An important aspect of processing the variables in this section involved assignment of legitimate skip codes, where relevant. That included (1) assignment of legitimate skip codes to variables in the entire module for respondents who were aged 18 or older, and (2) assignment of legitimate skip codes to youths' data based on routing logic within the youth mental health service utilization module. For example, if respondents reported that they did not stay overnight or longer in a hospital to receive mental health counseling in the past 12 months (YUHOSPYR = 2), all subsequent variables pertaining to mental health services in a hospital were assigned legitimate skip codes. That included the number of nights that respondents stayed in a hospital and the reasons that they were hospitalized the last time.

Although respondents in the youth experiences module who reported that they were not enrolled in school in the past 12 months were asked whether they were homeschooled during this period, the youth experiences variable pertaining to homeschooling (YEHMSLYR, corresponding to question YE09A) was not used to edit youth mental health service utilization variables pertaining to receipt of school-based mental health services. Only the youth experiences variable pertaining to school enrollment in the past 12 months (YEATNDYR, corresponding to question YE09) was used to edit these school-based service variables.

If respondents reported that they stayed overnight or longer in foster care or in a therapeutic foster care home in the past 12 months for emotional or behavioral problems, they were not asked whether they had ever been in foster care. Therefore, the edited variable pertaining to foster care in the lifetime (YUFCAREV) was assigned a code of 5 (Yes LOGICALLY ASSIGNED [from skip pattern]). This code of 5 indicated that it could be logically inferred that respondents had ever been in foster care because they reported being in foster care in the past 12 months.

Similarly, if the variable pertaining to foster care in the past 12 months (YUFCARYR) initially had a missing value (e.g., if respondents did not know or refused to report whether they stayed in foster care in the past 12 months) but respondents reported that they had never been in foster care (YUFCAREV = 2), it could be inferred that these respondents had not been in foster care in the past 12 months. In these situations, the edited variable YUFCARYR was assigned a final code of 4 (No LOGICALLY ASSIGNED). The remaining variables related to foster care in the past 12 months were assigned legitimate skip codes.

For each type or location of mental health treatment or counseling that respondents were asked about, they could report that they received services the last time at that particular location for any of the following reasons: (1) they thought about or tried to kill themselves, (2) they felt depressed, (3) they felt very afraid or tense, (4) they were breaking rules or "acting out," (5) they had eating problems, (6) they had trouble controlling their anger, (7) they had gotten into physical fights, (8) they had problems at home or in their families, (9) they had problems with their friends, (10) they had problems with people other than friends or family, (11) they had problems at school, or (12) some other reason. The reasons pertaining to trouble controlling anger through problems at school have been included since 2005 based on identification of commonly reported "other" reasons that respondents specified prior to 2005 for why they received treatment.

For each mental health service location where youths received services, information on these reasons for receiving services subsequently was captured as a discrete variable. For example, if respondents reported receiving mental health counseling from a pediatrician or family doctor, information about why they received counseling the last time was captured in the variables YUFDSUIC (suicidal), YUFDDEPR (depressed), YUFDFEAR (afraid and tense), YUFDBKRU (breaking rules), YUFDEATP (eating problems), YUFDANGR (anger), YUFDFITE (physical fights), YUFDFMLY (problems in family), YUFDFRND (problems with friends), YUFDOTPP (problems with people other than family or friends), YUFDSCHL (problems at school), and YUFDSOR (some other reason). The individual edited variables for these "enter all that apply" variables were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (see Section 2.4.4). No further editing was done if respondents endorsed every single reason on a list as pertaining to why they received mental health services at a given location in the past 12 months.

Codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to an entire list of variables if respondents did not know or refused to report why they received counseling at a specific location in the past 12 months; this applied as well to the items for the additional reasons why respondents received counseling that were added in 2005. If an entire list of reasons was blank but respondents had previously reported receiving services at a given location (e.g., if respondents broke off the interview), then the list of reasons for receiving services at that location retained a code of 98 (i.e., "blank").

For purposes of illustration, if youths reported in question YSU22 that they had received treatment or counseling from a pediatrician or other family doctor in the past 12 months (i.e., for emotional or behavioral problems that were not caused by alcohol or drugs), they were routed first to question YSU23 regarding the number of times they received treatment from a family doctor, and then they were routed to question YSU24, regarding the reason(s) for their visit; question YSU24 has been present in the module ever since the module was added to the survey.[121] As in prior years, YSU24 was an "enter all that apply" item, and individual variables were created corresponding to the specific reasons why respondents got treatment.

[121] Questions in this module for treatment that youths received in other settings were structured in the same manner as in this example for treatment from a family doctor. The changes to the questions for additional reasons why youths received treatment from a family doctor also applied to the other sources of treatment in this module.

Respondents could report up to six reasons in YSU24 for why they received treatment from a family doctor, including "some other reason." If respondents chose the "some other reason" option in question YSU24, they were routed next to question YSU24A, where they could choose the additional reasons for receiving treatment that were noted above, such as difficulty controlling anger; again, respondents were given the option in question YSU24A to report "some other reason." If respondents chose this "other" response category in YSU24A, they were asked to specify the *most important* other reason why they did got treatment in this location; this request for the most important other reason was a noteworthy change relative to prior years (see below).

Based on the logic noted above, the sixth response option in question YSU24 (some other reason) was principally considered to be a "toggle" to question YSU24A. Therefore, a separate "some other reason" variable was not created to correspond to the last response category in YSU24.

The edits described below have been implemented since 2005 to take into account the items on additional reasons why youths received treatment (e.g., YSU24A).

- If respondents chose the sixth response option in YSU24, any response that was entered from the YSU24A series was coded as 1, and anything that respondents did not choose from the YSU24A list was coded as 6.
- If respondents chose a response from YSU24 but did not choose the sixth response category in YSU24, the variables corresponding to the response categories in YSU24A (i.e., YUFDANGR through YUFDSOR) all were given a code of 6 (Response not entered), rather than being assigned "legitimate skip" codes. That is, YSU24 and YSU24A were considered together to be one big series of reasons.
- If respondents chose the sixth category in question YSU24, chose at least one reason from YSU24A, but did not choose category 7 in YSU24A (some other reason), the edited variable YUFDSOR (some other reason for receiving treatment) was assigned a code of 6. That is, it was inferred in this situation that the list of specific reasons in YSU24A was adequate for capturing why respondents saw a family doctor about emotional or behavioral problems in the past 12 months. For example, if a respondent chose category 6 in YSU24 and then chose only category 3 in YSU24A ("You had problems at home or in your family"), it would be reasonable to infer that this response in YSU24A was the only other reason why the respondent visited a family doctor about emotional or behavioral problems.
- If respondents chose the sixth response category in question YSU24, it was possible for them to answer YSU24A as "don't know" or "refused" (i.e., did not know or refused to report what the other reasons were). When this occurred, the "some other reason" variable YUFDSOR was set to 1 (Response entered) in order to retain information that the respondent chose "some other reasons" somewhere in the series. Remaining variables corresponding to the YSU24A series retained codes of 94 ("don't know") or 97 ("refused").

- If respondents answered question YSU24 as "don't know" or "refused," question YSU24A was skipped. Therefore, the relevant code of 94 or 97 was propagated onto the variables corresponding to the YSU24A list.
- If YUFDSOR had a value of 6 (see above), the "OTHER, Specify" variable YUFDIMPR (i.e., the most important other reason why the respondent received treatment from a family doctor) was assigned a legitimate skip code. If YUFDSOR had a refusal code, that refusal was propagated onto YUFDIMPR.
- If YUFDSOR had a code of 1 when the respondent answered "don't know" or "refused" to the YSU24a series, the "OTHER, Specify" variable YUFDIMPR (which has been three digits in length since 2005) retained a code of 998 (blank).

Consistent with general editing procedures, if respondents reported a reason that corresponded to a reason in the lists for YSU24/YSU24A, that reason was logically inferred to have been chosen in the relevant edited variable. Suppose, for example, that the most important other reason that a respondent reported for receiving treatment from a family doctor indicated that the respondent was breaking rules or "acting out." If the respondent had not chosen this response in YSU24, the edited variable YUFDBKRU was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). Similarly, if the respondent specified that a problem in the respondent's family was the most important other reason for receiving treatment from a family doctor and the respondent had not chosen this response in YSU24A, the edited variable YUFDFMLY was assigned a code of 3. Conversely, if respondents did not report "some other reason" why they received treatment in the past 12 months from a family doctor (edited variable YUFDSOR = 6, corresponding to response category 6 in question YSU24 not being chosen or response category 6 in question YSU24A not being chosen), legitimate skip codes were assigned to the edited "OTHER, Specify" variable YUFDIMPR (corresponding to question YSU24SP).

As noted previously, the "OTHER, Specify" items in this module underwent an important change in 2005. Prior to 2005, youths were asked to specify "the other reason" why they received services in a particular location or from a given type of provider. Some respondents in these prior years gave a considerable amount of information in the space that was allotted to them to specify their other reason(s) for receiving services. Often, multiple reasons were reported. Therefore, prior to 2005, up to five separate "OTHER, Specify" codes were assigned for a given treatment location or provider based on respondents' explanations regarding why they received services.

Because youths since 2005 have been asked to specify the "most important" other reason why they received services in a particular location, only one reason was captured in the "OTHER, Specify" variables. If respondents specified more than one reason as the "most important" other reason why they received treatment in a given location, only the first reason that respondents specified was coded as a general rule. The exception to this rule was that any reports of respondents thinking about or trying to kill themselves were given precedence in coding, regardless of whether this reason was specified first. In addition, the change in the "OTHER, Specify" variables for this module was used as an opportunity to revise the "OTHER, Specify" codes to regroup related reasons together, such as reasons that indicated that respondents had a diagnosed condition (e.g., attention-deficit/hyperactivity disorder, or ADHD). For these reasons, the "OTHER, Specify" data from 2005 onward are not comparable with data prior to 2005.

In a relatively rare number of situations, youths denied receiving mental health services as part of an "OTHER, Specify" response. In these situations, the "OTHER, Specify" response was assigned a bad data code. Data were retained that indicated that the youths received mental health services in a given location in the past 12 months.

Respondents could report that the number of nights they stayed overnight in a hospital or residential treatment program in the past 12 months (or the sum of the two, if respondents reported staying in both settings) was greater than or equal to 365 nights. In these situations, no editing was done to the data. If respondents reported spending 366 nights in an inpatient/residential setting or in foster care for treatment or counseling for emotional or behavioral problems, however, the number of nights in the corresponding edited variable was "trimmed" back to 365. For example, if respondents reported staying overnight in a hospital for 366 nights for treatment or counseling for emotional or behavioral problems, the corresponding edited variable YUHOSPNM was set to a value of 365.

7.4.17 Consumption of Alcohol Module

The consumption of alcohol module covered a variety of topics related to respondents' use of alcohol. Administration of questions in this module was limited to respondents who had ever used alcohol, and a subset of questions was limited to respondents who reported alcohol use in the past 30 days. Respondents were asked questions about the following topics:

- number of drinks consumed on the last occasion of alcohol use in the past 30 days;
- the last occasion of alcohol use in the past 30 days by persons under the legal drinking age of 21, including who respondents were with, where they drank, and how they obtained the alcohol;
- use of other drugs in combination with alcohol (i.e., at the same time or within a couple of hours of using alcohol) in the past 30 days; and
- history of binge alcohol use (i.e., consumption of five or more drinks on a single occasion for males or four or more drinks on a single occasion for females).

This section documents the editing procedures for the consumption of alcohol module.

7.4.17.1 Skip Logic Issues in the Consumption of Alcohol Module

One of the important aspects of the processing of variables in the consumption of alcohol module consisted of assigning codes of 91, 93, and 99 to variables that had been skipped because the questions did not apply. In particular, certain questions may not apply to respondents for multiple reasons, such as if they never used alcohol or if they were male.

Therefore, Table B.26 in Appendix B discusses the prioritization of skip logic edits in this module. For example, if respondents had never used alcohol, variables in the consumption of alcohol module were assigned a code of 91, 991, or 9991 (NEVER USED ALCOHOL), regardless of whether questions might not be applicable for other reasons. In particular, male respondents who had never used alcohol were assigned codes of 91, 991, or 9991 in the questions about females' history of consuming four or more drinks on a single occasion, even

though males were not eligible for these questions because of their gender. However, this edit kept a constant number of cases coded as indicating nonuse of alcohol for most variables in the module.

7.4.17.2 Edit Issues for the 30-Day Consumption of Alcohol Variables

Other than implementation of edits based on the skip logic (see Table B.26), only limited editing was done for the variables pertaining to the number of drinks that respondents consumed the last time they drank in the past 30 days (question CA01; edited variable CADRLAST) and the variables pertaining to alcohol use in the past 30 days among persons aged 12 to 20 (questions CA02A through CA07SP). In particular, no editing was done to CABUYFRE (corresponding to question CA03), CAGVMONY (corresponding to question CA04), CABUYWHO (corresponding to question CA05A), or CABPLACE (corresponding to question CA05B) if respondents reported in question CA04 that they gave money to someone else who bought their last alcoholic beverage for the respondents, but then they reported in CAFRESP (corresponding to CA07SP) that the "other" way that they got the last alcoholic beverage they drank was by purchasing it, such as with fake identification. The rationale for not editing these variables was that there may not be enough detail from what respondents specified to know how these variables should be edited. For example, question CA04 about buying alcohol was asked only if respondents reported in question CA03 that they paid for the last alcohol that they drank. Even if the respondent specified buying alcohol in CAFRESP, however, it would not be known whether the respondent bought the alcohol with his or her own money or whether the alcohol was purchased with someone else's money.

In question CA07 (edited variable CAFREWHO), respondents were asked to give one response for how they got the last alcoholic beverage that they drank, including the option to report that they got it "some other way." Respondents who reported that they got the alcohol some other way were asked to specify what this other way was (edited variable CAFRESP). If a response was given in CAFRESP that corresponded to one of the existing response options in question CA07, then CAFREWHO was assigned the relevant response plus a value of 10. For example, if CAFRESP indicated that respondents got their last alcoholic beverage from a nonrelative who was 21 or older who gave the alcohol to the respondent (i.e., equivalent to CA07 = 3), then CAFREWHO was assigned a code of 13.

Questions CA02B (where underage respondents drank alcohol the last time) and CA09 (the specific drugs that respondents of any age used at the same time or within a couple of hours of when they last drank alcohol) were "enter all that apply" questions. As in other modules, the responses were captured as discrete variables. The individual edited variables for these "enter all that apply" variables were coded as 1 or 6 to indicate that the response was entered or not entered, respectively (see Section 2.4.4).

For the edited variables corresponding to question CA02B, if respondents specified some other place where they drank alcohol the last time that corresponded to one of the existing response options in question CA02B, and they had not chosen that response, then the corresponding edited variable was assigned a code of 3, where 3 = Response entered LOGICALLY ASSIGNED. For example, if respondents reported that they drank alcohol in someone else's home the last time they drank but they had not chosen this response in question

CA02B, then the edited variable CADROTHM was assigned a code of 3. In addition, codes of 94 and 97 (for "don't know" and "refused," respectively) were assigned to the entire list of variables corresponding to question CA02B if respondents did not know or refused to report where they were the last time they drank alcohol.

For the edited variables corresponding to question CA09, codes of 94 or 97 were retained only for those drugs that respondents used in the past 30 days (or were potentially past month users, such as if a core recency variable had been coded as 8, meaning that the respondent used the drug at some point in the past 12 months, which could include within the past 30 days). Otherwise, codes of 94 or 97 in the individual drug variables corresponding to question CA09 were replaced with codes of 89 (LEGITIMATE SKIP Logically assigned) if respondents used some drug in the past 30 days but not the drug that corresponded to the variable of interest. Suppose, for example, that respondents used marijuana and inhalants in the past 30 days and did not use any other illicit drugs in that period. If question CA08 was answered as "yes" and CA09 was refused, the edited variables CADRKMRJ and CADRKCOC retained codes of 97, and the remaining drug variables were assigned codes of 89.

Special consideration also was given to the editing of the variables CADRKCOC (last use of alcohol in combination with cocaine), CADRKHER (last use of alcohol in combination with heroin), CADRKSTM (last use of alcohol in combination with prescription stimulants), and CADRKMTH (last use of alcohol in combination with methamphetamine). In particular, respondents who answered question CA08 as "yes" would have the response option "Cocaine or 'crack'" (captured in CADRKCOC) available to them if they previously reported past month use of cocaine in any of the following places: (1) in the core cocaine module, (2) in the core crack cocaine module, or (3) in the noncore special drugs module (i.e., use of cocaine with a needle in the past 30 days). Similarly, respondents who answered question CA08 as "yes" could choose the response option "Heroin" (captured in CADRKHER) if they reported past month use of heroin in the core heroin module or if they reported past month sniffing, smoking, or injection of heroin in the noncore special drugs module. The response option "Prescription stimulants" (captured in CADRKSTM) was available to respondents if they reported nonmedical use of stimulants in the past 30 days in the core stimulants module or if they reported that they used stimulants other than methamphetamine with a needle in the past 30 days. For methamphetamine, the response option "Methamphetamine" (captured in CADRKMTH) was available to respondents who indicated past month methamphetamine use in the core stimulants module or who indicated past month use in the noncore special drugs module, provided that these additional methamphetamine users from the noncore special drugs module did not indicate in question SD17A1 that their earlier report from the core that they never used methamphetamine was actually the correct response (see Section 6.2.6).

Consequently, respondents would have the response options for CADRKCOC, CADRKHER, CADRKSTM, or CADRKMTH available to them if they reported past month use (or nonmedical use) of the relevant drug(s) in the noncore special drugs module, even if they had not reported past month use in the corresponding core module(s). In most situations, no further editing was done to these data, even though the corresponding core data did not indicate use in the past 30 days. In particular, no editing was done to data in CADRKMTH if respondents reported in the core stimulants module that they never used methamphetamine, but they reported past month use of methamphetamine in the special drugs module and they also reported in the

follow-up question SD17A2 that they "made a mistake" (or a similar other reason for not reporting methamphetamine use) when they were answering the core question about methamphetamine.

The one exception to this editing principle concerned editing of CADRKSTM. Respondents who reported nonmedical use only of methamphetamine and of "some other stimulant" in the core stimulants module were routed to question ST09 (for the most recent nonmedical use of any stimulant) and to question ST19 (for the most recent use of methamphetamine). If respondents reported nonmedical use of stimulants in the past 30 days in question ST09 but they were logically inferred to be lifetime nonmedical users of only methamphetamine (i.e., the only "other" stimulant that they specified using nonmedically was methamphetamine), then any data in the response option corresponding to CADRKSTM were overwritten with a code of 89 as part of the data editing. That is, it was logically inferred that these respondents should not have been asked about their use of *prescription* stimulants while they were drinking alcohol or within a couple of hours of drinking alcohol in the past 30 days.

7.4.17.3 Edit Issues for the Binge Alcohol History Variables

Questions CA10 through CA14D captured information on respondents' history of having consumed five or more drinks on a single occasion (regardless of gender) or four or more drinks on a single occasion (for females). If respondents were lifetime alcohol users and other data from the core alcohol module or from CADRLAST did not indicate that respondents had consumed five or more drinks on a single occasion in the past 30 days, then they were asked in question CA10 (edited variable CABNGEV) whether they had ever had five or more drinks on a single occasion in their lifetime. Those respondents who had ever consumed five or more drinks on a single occasion (or had reported this in the past 30 days in the core alcohol module or in CADRLAST) were asked in question CA11 (edited variable CABNGAGE) to report how old they were the first time when they first had five or more drinks. If CABNGAGE was within 1 year of respondents' current age, they were routed to questions CA11A through CA11D (edited variables CABNGYFU and CABNGMFU) to report the year and month when they first had five or more drinks on a single occasion.

Similarly, female respondents were asked question CA12 (edited variable CA4FDEV) if they were lifetime alcohol users and other data from the core alcohol module or from CADRLAST did not indicate that they had consumed four or more drinks on a single occasion in the past 30 days, and CABNGEV did not indicate that they had ever consumed five or more drinks on a single occasion. In addition, question CA13 (edited variable CA4FDDYS) asked female respondents who used alcohol in the past 30 days to report the number of days they had four or more drinks on a single occasion in the past 30 days. Question CA14 (edited variable CA4FDAGE) asked female respondents to report how old they were the first time they had four or more drinks on a single occasion. If CA4FDAGE was within 1 year of female respondents' current age, they were routed to questions CA14A through CA14D (edited variables CA4FDYFU and CA4FDMFU) to report the year and month when they first had four or more drinks on a single occasion.

As noted previously, the routing logic for these questions took into account respondents' answers in the core alcohol module. Specifically, the computer-assisted interviewing (CAI)

program created a flag variable called FOURORMOREFLAG. The default value for this flag was 0. The flag was coded as 2 if there was some indication in the alcohol module that respondents had consumed five or more drinks on a single occasion in the past 30 days. Otherwise, if question AL07 indicated that respondents usually had four drinks on those days when they drank in the past 30 days, then this flag was coded as 1.

The edited variable corresponding to this flag was named ED4FLAG. If the variables NODR30A (corresponding to question AL07) and DR5DAY (corresponding to question AL08) indicated consumption of five or more drinks on a single occasion but ED4FLAG did not already have a value of 2, then ED4FLAG was set to a logically assigned value of 12. Otherwise, if NODR30A indicated usual consumption of four drinks in the past 30 days and ED4FLAG did not already have a value of 1, then ED4FLAG was set to a value of 12. If NODR30A and DR5DAY had been set to bad data as a result of edits in the core alcohol module (see Section 6.2.4.7 and Table B.7 in Appendix B), then ED4FLAG also was set to bad data. In turn, any questions that had been answered in CA11 through CA14D were set to bad data.

Revised routing logic for variables related to binge alcohol history has been in place since 2007 based on patterns that were identified during review of the edited data in the 2006 survey. Specifically, the routing logic was revised for question CA11 (edited variable CABNGAGE) to take into account FOURORMOREFLAG (edited variable ED4FLAG). In 2006, the logic took into account respondents' original answers about consumption of five or more drinks on a single occasion in the core alcohol question AL08. However, respondents in 2006 who changed their answer to indicate that they had five or more drinks on 0 days in the past 30 days in response to a consistency check between the 30-day frequency of any alcohol use and the frequency of binge alcohol use in that period were incorrectly asked question CA11 based on their original answer in AL08. The FOURORMOREFLAG variable reflected this consistency resolution between the overall 30-day frequency and the frequency of consumption of five or more drinks, such that these respondents have been correctly skipped out of question CA11 since 2007. Consequently, no data in CABNGAGE were overwritten with codes of 989 (LEGITIMATE SKIP Logically assigned) to indicate that respondents should not have been asked CA11.

In addition, updated routing logic for females for questions CA13 (edited variable CA4FDDYS) and CA14 (edited variable CA4FDAGE) has been in place since 2007 to include FOURORMOREFLAG in addition to questions CA01 (edited variable CADRLAST) and CA12 (edited variable CA4FDEV). This corrected for a logic error in 2006 in which female respondents were routed to CA13 and CA14 only if they answered question CA12 as "yes" or reported in question CA01 (CADRLAST) that they had four or more drinks the last time they drank alcohol in the past 30 days. Those female respondents in 2006 who were skipped out of question CA12 based on the flag variable from the core alcohol module (i.e., ED4FLAG) or who reported consumption of five or more drinks in question CA10 (CABNGEV) were not asked subsequent questions about their consumption of four or more drinks, unless they reported in question CA01 that they last had four or more drinks. Because of this change in 2007, no data in CA4FDAGE were assigned a code of 990 (NOT ASKED THE QUESTION Logically assigned) due to female respondents being incorrectly skipped out of question CA12. Similarly, no codes of 90 (which had the same meaning as that for 990) were assigned to CA4FDDYS in 2007 if female respondents had used alcohol in the past 30 days.

Table B.27 discusses additional edits associated with the binge alcohol history variables in the consumption of alcohol module. The first set of edits pertains to the consistency checks between the age when respondents reported that they first had five or more drinks (or four or more drinks for females) in a single occasion, and the age at initiation based on the year and month when respondents reported that they first engaged in this behavior.

Table B.27 then presents edits for these binge alcohol variables relative to respondents' ages and to corresponding core alcohol use data. For example, if respondents reported in question CA11 that they first had five or more drinks on a single occasion at an age that was earlier than what they reported for when they first had a drink of any alcohol (from the core alcohol use variable ALCTRY), then CABNGAGE was set to bad data. Any nonmissing values in CABNGYFU and CABNGMFU also were replaced with codes for bad data. Similarly, suppose respondents first used any alcohol at their current age. If they had a lifetime history of consumption of five or more drinks on a single occasion (including any indication of this behavior in the past 30 days) but CA11 had a missing value (e.g., "don't know" or "refused"), then it could be logically inferred in CABNGAGE that the first episode of binge alcohol use occurred at the respondents' current age.

For female respondents, the data pertaining to consumption of five or more drinks on a single occasion and consumption of four or more drinks on a single occasion also were edited for consistency with one another. For example, if CABNGAGE (from question CA11) indicated consumption of five or more drinks at an age that was earlier than what respondents reported in CA14, then CA4FDAGE was assigned the earlier age from CABNGAGE.

In addition, "editing indicator" (EI) variables were created to indicate when nonmissing values were logically assigned to the binge alcohol variables in the consumption of alcohol module. The default value for these EI variables was 1 (i.e., questionnaire data). The EI variables were assigned a value of 2 (Logically assigned data) if a logical inference was made that involved assignment of a nonmissing value. In the above example, if CA4FDAGE was assigned an earlier age from CABNGAGE, then the corresponding variable EI4FDAGE was assigned a value of 2.

In keeping with the general principle of not using noncore data to edit core data, no editing was done to core data when a respondent's reported first consumption of five or more drinks on a single occasion (or four or more drinks for females) suggested more recent use of alcohol than was indicated in the core recency variable ALCREC. For example, ALCREC may not indicate use in the past month or past year, but respondents could report that they first had five or more drinks or four or more drinks on a single occasion at their current age. Although these data from the consumption of alcohol would suggest that these respondents used alcohol in the past 12 months, no editing was done to ALCREC based on these data.

8. Editing and Imputation for the NSDUH Roster Variables

8.1 Introduction

This chapter describes the techniques used to edit and impute variables associated with the household roster for the 2012 National Survey on Drug Use and Health (NSDUH). The variables described in this chapter can be divided into three groups:

- respondent-level detailed roster variables,
- roster-derived household composition variables, and
- proxy variables.

The respondent-level detailed roster variables included the age, gender, and relationship to the respondent for each household member. The introductory question for the household roster portion of the questionnaire (QD54) was interviewer administered. This question asked the respondent how many persons lived in the household. The computer-assisted interviewing (CAI) instrument was set up to be able to collect data on up to 25 household members. If only one person lived in the household or the respondent did not know or refused to answer, then the household composition (roster) section was skipped. Otherwise, the respondent was asked questions about the age, gender, and relationship to the respondent of every member of the household, starting with the household's oldest member and including the respondent.

If a pair of respondents was selected in a household, the interviewer indicated which member of a respondent's household roster corresponded to the other selected pair member. The roster entry for the respondent was referred to as the "self" entry. In effect, the respondent completed a grid with the number of rows corresponding to the value entered in QD54. Table 8.1 shows an example grid where the number of persons in the household is four. In this example, the roster of the respondent is shown, and the indicator variable shows that the respondent's son was selected as the other pair member. The possible relationship codes and specific relationship details between pair members are listed in Table 8.2.

Table 8.1 Roster Grid Example Where Number of Persons in Household (QD54) Equals 4

Person #	Relationship to Respondent	Age in Years	Other Member Selected for Pair[1]
1	Self	44	0 (No [Impossible])
2	Husband	42	0 (No)
3	Son	16	1 (Yes)
4	Boarder/Roomer	16	0 (No)

[1] This indicator variable applied only to respondents who were part of a pair selection. The other member selected could not have been the self because respondents were not interviewed twice. The other member selected was the roster member who had a value of "1" for this variable.

Table 8.2 Roster Relationship Codes

Relationship Code #	Relationship to Respondent	Details about Relationship
1	Self	
2	Parent	Biological, Step, Adoptive, or Foster
3	Child	Biological, Step, Adoptive, or Foster
4	Sibling	Full, Half, Step, Adoptive, or Foster
5	Spouse	
6	Unmarried Partner	
7	Housemate or Roommate	
8	Child-in-Law	
9	Grandchild	
10	Parent-in-Law	
11	Grandparent	
12	Boarder or Roomer	
13	Other Relative	
14	Other Nonrelative	

The second group of variables, the roster-derived household composition variables, was derived from the respondent-level detailed roster variables. These were mostly count variables reporting the number of individuals in the roster with various characteristics. These variables are listed in Table 8.3, which also shows that some of these variables underwent imputation. Among the three groups of variables described in this chapter, only the roster-derived household composition variables underwent imputation of any kind.

The third group of variables, the proxy variables, allowed for the selection and identification of a relative of the respondent who lived in the respondent's household (according to the household roster), who was aged 18 years or older and who answered the health insurance coverage and income questions for the respondent. The edited versions of these variables and the questions to which they map are shown in Table 8.4.

Table 8.3 Roster-Derived Household Composition Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Total number of rostered persons	TOTPEOP	IRHHSIZE
Number of persons in household aged 17 or younger	KID17	IRKID17
Number of persons in household aged 65 or older	HH65	IRHH65
Indicator of whether the respondent had family members in household	FAMSKIP	IRFAMSKP
Number of respondent's family members in household (includes foster relationships)	FMLYSIZE	IRFMLYSZ

Table 8.3 Roster-Derived Household Composition Variables (continued)

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of respondent's family members in household aged 17 or younger (includes foster relationships)	KIDFMLY	IRKDFMLY
Number of respondent's family members in household (excludes foster relationships)	FAMSIZE	IRFAMSZE
Number of respondent's family members in household aged 17 or younger (excludes foster relationships)	KIDFAMSZ	IRKIDFAM
Number of respondent's children in household aged 2 or younger	NRBABIES	N/A
Number of respondent's children in household aged 3 to 5 years old	NRPRESCH	N/A
Number of respondent's children in household aged 6 to 11 years old	NRYUNGCH	N/A
Number of respondent's children in household aged 12 to 17 years old	NRTEENS	N/A
Number of respondent's children in household aged 17 or younger	NRCH0_17	N/A
Number of respondent's children in household aged 18 to 20 years old	NROLDRCH	N/A
Number of respondent's children in household aged 21 or older	NROLDCH	N/A
Number of roommates/housemates in household	NROOMATE	N/A
Indicator of presence of mother in household (12- to 17-year-olds)[1]	IMOTHER	N/A
Indicator of presence of father in household (12- to 17-year-olds)[1]	IFATHER	N/A
Indicator of presence of foster child in household (12- to 14-year-olds)[2]	FSTRCHLD	N/A

[1] The IMOTHER and IFATHER indicators were not 0/1 indicators because levels were provided for "unknown" and "18 or older."
[2] This variable was required for the creation of the POVERTY variable for the 2003-2005 survey years.

Table 8.4 Proxy Variables

Unedited Variable	Text of Survey Question Associated with Unedited Variable	Edited Variable
QP01	Is there anyone else who lives here who is 18 or older who would be better able to give me the correct information about your health insurance coverage and the kinds of income you receive?	PRXABLE2
QP02	Who is the person you think can help us get the correct information for these questions?	PRXRELAT
QP03	Is your [QP02 fill] available right now?	PRXHOME2
QP04	Would you ask your [QP02 fill] to join us to help with these last questions about health insurance and income?	PRXJOIN2
HASJOIN	Has the person's [QP02 fill] joined R?	PRXYANS2

8.2 Editing the Respondent-Level Detailed Roster Variables

This section describes the methods used to create edited versions of the respondent-level detailed roster variables: ROSAGE1-ROSAGE25, ROSSEX1-ROSSEX25, ROSRLT1-ROSRLT25, ROSMSL1-ROSMSL25, PRNTYP1-PRNTYP25, SIBTYP1-SIBTYP25, CHDTYP1-CHDTYP25, and TWNTYP1-TWNTYP25. These variables describe up to 25 members of the household. The editing procedures for the respondent-level detailed roster variables began with consistency checks included in the Blaise program code, which were implemented to reduce the amount of editing required at the data processing stage. The consistency checks in the questionnaire were supplemented with other edits involving the respondent-level detailed roster variables outside the CAI instrument. These involved resolving cases where it was unclear which roster member was the self and cases where relationship codes were impossible (or very unlikely) given the age and gender in relation to the self.

Section 8.2.1 describes the consistency checks programmed into the questionnaire. Section 8.2.2 describes the creation of a roster-level dataset for further processing. Section 8.2.3 describes roster edits involving the self. Section 8.2.4 describes roster edits for other household members, after the self has been established. Finally, Section 8.2.5 describes the creation of the final edited respondent-level detailed roster variables.

8.2.1 Roster Consistency Checks

Two types of consistency checks were employed in the CAI instrument for the household roster section of the questionnaire. These checks (1) compared the roster entry corresponding to the respondent with previously entered questionnaire information and (2) compared a roster entry against other roster entries for internal consistency. With the exception of the check against the previously entered respondent's gender, the interviewer could override the consistency checks and explain why the response given was correct. Interviewers' explanations for overrides to consistency checks and evaluations of their validity are provided in Appendix G.

8.2.1.1 Comparisons with Previously Entered Questionnaire Information

Gender and age were the two consistency checks built into the household roster section of the CAI instrument that compared the roster entry with the previously entered questionnaire information. The check for gender was added in 2001 and was triggered if the respondent in the household roster entered a gender that was different from the one previously recorded in the interview (question QD01). If the gender did not match, the interviewer was required to change either the roster entry or the gender that had been entered at the beginning of the interview.

The consistency check for age was added in 2002 to compare the respondent’s age in the roster with the age previously entered in the questionnaire (the Blaise variable CURNTAGE). The interviewer could either change the respondent's age entered in the roster or override the consistency check. If the interviewer chose to override the consistency check, then he or she provided an explanation as to why the roster age did not match CURNTAGE. Explanations given by the interviewer for overriding this particular consistency check were carefully reviewed. Interviewers' explanations for overrides to consistency checks and evaluations of their validity are provided in Appendix G. In rare cases, the final value for age (AGE) was set to the

age of the self in the household roster (the "roster age") based on these explanations as well as other evidence. Additional details about how roster age was used for creating AGE are described in Chapter 4. Strategies for the more common situation, where the original value for AGE was not set to the roster age, are discussed in Section 8.2.3.

8.2.1.2 Internal Consistency Checks

Since the 2002 survey, internal consistency checks have been implemented in the CAI instrument to compare one roster entry with another. These checks were triggered if any of the following conditions occurred:

1. The interviewer reported that the respondent had more than one spouse or unmarried partner or reported a spouse and an unmarried partner.
2. The interviewer reported that a household member was a parent or grandparent of the respondent and the respondent was older than the household member.
3. The interviewer reported that a household member was a child or grandchild of the respondent and the respondent was younger than the household member.
4. The interviewer reported that a household member was a spouse or an unmarried partner of the respondent and the household member was 16 years old or younger.
5. The interviewer reported that the respondent had a spouse or unmarried partner and the respondent was 16 years old or younger.
6. The interviewer reported that the respondent was either a child-in-law or a parent-in-law and the respondent was 16 years old or younger.
7. The interviewer reported that a household member was a child-in-law of the respondent and the household member was the same age or older than the respondent.
8. The interviewer reported that a household member was a parent-in-law of the respondent and the household member was the same age or younger than the respondent.
9. The interviewer reported that a household member was a biological parent of the respondent and the household member was less than 13 years older than the respondent.
10. The interviewer reported that a household member was a biological child of the respondent and the household member was less than 13 years younger than the respondent.
11. The interviewer reported that a household member was a biological sibling of the respondent and the household member was more than 24 years older or younger than the respondent.
12. The interviewer reported that a household member was a grandparent or grandchild of the respondent and the age difference was less than 30 years.

In most cases, if a consistency check was triggered, the interviewer changed either an age code or a relationship code in the roster to a more appropriate value. Any edit that was invoked

because of an override to a consistency check was carefully scrutinized during the data processing stage. The relevant household roster, as well as the explanation given by the interviewer for the override, was carefully examined to determine whether the override was legitimate. If the override was deemed legitimate (e.g., a father marries a woman, listed as [step] mother, who is younger than the respondent), the original answer was allowed to remain and no edit was applied. If the interviewer's explanation was not considered legitimate, then an edit was applied. More details about roster edits are provided in Section 8.2.4. Explanations given by the interviewers for the overrides and evaluations of their legitimacy are provided in Appendix G.

8.2.2 Roster-Level Dataset

To facilitate processing of the roster variables, a roster-level dataset was created in which the number of records per respondent was given by the household size in question QD54. For example, if a respondent indicated a household of three consisting of himself or herself, a mother, and a father, then there would be three records on the dataset associated with this respondent: one for the self, one for the mother, and one for the father. Even if the respondent did not start or complete the roster questions, records were created for each household member.

8.2.3 Roster Edits Involving the Self

The Blaise program code required the interviewer to identify exactly one self and a corresponding age and gender in the household roster. Moreover, the interviewer was required to confirm with the respondent that the respondent was in fact the identified self. Because the check involving gender was not allowed to be overridden, the gender for the self in the roster always matched QD01, which was equivalent to IRSEX (see Chapter 4). However, it was possible to have problems matching AGE (see Chapter 4 for a description of the methodology used to create AGE) with the age of the self in the roster, despite the consistency check comparing the respondent's roster age against CURNTAGE.

The interviewer was able to override the consistency check for age of the self for one of two reasons: (1) the self was misidentified and another roster member was the true self but the interviewer decided not to change the entries; or (2) the interviewer correctly identified the self but indicated that the correct age for the respondent was different than CURNTAGE, and other evidence did not support this claim (AGE was not set to the roster age, as discussed in Section 8.2.1.1). In the case of a misidentified self, a second roster member in the household was selected whose gender matched IRSEX and whose age was within 1 year of AGE. The second roster member who replaced the original self had an age and gender that matched IRSEX and AGE, respectively.

If the consistency check was overridden, a misidentified self was diagnosed if (1) the roster age of the self differed from AGE by more than 1 year, and (2) another roster member of the same gender as QD01 (and IRSEX) had a roster age within 1 year of AGE.[122] If a misidentified self was diagnosed, it was assumed that the interviewer used the roster member identified as the self, rather than the respondent, as the point of reference. Using the example shown in Table 8.1, if the respondent's son was used as the reference point, the relationship for

[122] A 1-year difference was allowed because the respondent's age might have changed during the interview. In this instance, the values of AGE and CURNTAGE may have differed by 1 year.

the respondent became "mother" instead of "self," and the "husband" became "father." Under these circumstances, the self code was set to missing, and the respondent's roster entries did not include a self. The remaining relationship codes in the roster also were set to missing. In some cases, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point.

8.2.3.1 Original Self Misidentified: Identifying the Real Self

If the self was misidentified in the roster, an attempt was made to identify a self among the roster members corresponding to the respondent. A roster member was selected as the self under one of two possible circumstances: (1) the roster member's age, gender, and relationship data were missing; or (2) the roster member was of the respondent's gender and was within 1 year of the respondent in age. If more than one roster member met the above criteria, the roster members who met the criteria but were not assigned the self code, were given a bad data code; that is, the original relationship code would no longer make sense because the reference person had been changed.

8.2.3.2 Original Self Misidentified: Salvaging Relationship Codes

As stated earlier, if the self was misidentified, all other relationship codes were set to missing because the reference person was someone other than the respondent. In some cases, however, the original relationship codes were salvaged, depending upon the roster member who was used as a reference point. Relationship codes were salvaged under the following circumstances:

1. If the reference person was the respondent's sibling, the roster member listed as "self" was actually a sibling, and all other relationship codes were salvaged. (Generally, relationships between the respondent and other household members would be the same with a sibling. For example, the respondent's parents are also the respondent's sibling's parents.)
2. If the reference person was the respondent's spouse or unmarried partner, the roster member listed as "self" was actually a spouse or unmarried partner, and the children relationship codes were salvaged.
3. If all the roster members other than the misidentified self were either roommates, boarders, or other nonrelatives, then the reference person was the respondent's roommate, boarder, or other nonrelative. All other relationship codes were salvaged.

8.2.4 Roster Edits for Other Household Members

Relationship codes were edited if the relationship of the roster member to the self was logically impossible based on age and gender. Edits of household roster ages, genders, and/or relationship codes were performed that either changed the reported value to another value or changed the reported value to bad data. It is important to note that in some cases, two members were selected in a household, which greatly increased the ability to edit the roster for those respondents. Some edits were associated with consistency checks, and interviewers' explanations for overrides to these consistency checks were carefully examined to assess the legitimacy of the override as explained in Section 8.2.1. Some edits were "automatic" in the programming code,

which meant that the interviewer was assumed to have been incorrect when the override was implemented. These edits were undone if the interviewer's explanation for the override was considered legitimate. In other situations, the default strategy was to assume that the override of the consistency check was correct and, therefore, that the edit was applied only if the interviewer's explanation appeared incorrect.

8.2.4.1 Edits to Roster Age, Gender, and Relationship Codes: Changes to Different Values (Correct Reference Person)

The following edits were performed on the roster age, gender, and relationship code values when the recorded age, gender, and/or relationship code was either missing or internally inconsistent and replaced by internally consistent values. When typing on a computer keyboard, it was possible for a double-digit age to have been entered as a single-digit age ("5" instead of "55"), or vice versa ("55" instead of "5"). If the relationship code still was believable, even with the incorrectly entered age (e.g., "other relative"), this type of error was difficult to detect because no inconsistency check was triggered. On the other hand, if an age entered this way triggered one of the consistency checks discussed in Section 8.2.1.2, the interviewer had an opportunity to correct the entry error. On those occasions where the age did not trigger a consistency check, detection of the error was still possible among selected pairs by examining the roster entries of the other pair member. For example, if one pair member had an x-year-old and no xx-year-olds, and the other had an xx-year-old and no x-year-old, where x denoted a single-digit number, it was highly probable that an error had occurred. By comparing the number of children younger than 12 years old in each roster with the number of children on the screener roster, it was apparent how a correction should be made. In this instance, the incorrectly entered age was replaced with the value given by the pair member whose roster age and screener age agreed.

1. If two members were selected in a household, the roster age for the other member selected was commonly not the same as the questionnaire-edited age (AGE, defined in Chapter 4) of the other pair member. In this case, the roster age for the other member selected was changed to this questionnaire-edited age value.
2. If two members were selected in a household, the gender that one member selected for the other on the household roster was often not the same as the gender (IRSEX, defined in Chapter 4) reported by that other pair member in his or her interview. In this case, the roster gender was changed to match the gender value the other pair member reported in his or her interview.
3. In previous survey years, the relationship codes for grandchild (9) and grandparent (11) were commonly confused. The following edit, which was used in previous survey years, was maintained in case of overrides: If the age of the respondent was at least 20 years older than that of the roster member, but the roster member was identified as a grandparent, the relationship code was changed to grandchild. Conversely, if the age of the respondent was at least 20 years younger than that of the roster member, but the roster member was identified as a grandchild, then the relationship code was changed to grandparent.

8.2.4.2 Edits to Relationship Codes: Changes to Missing Codes

The following edits were performed on the roster relationship code values, where the relationship code given was internally inconsistent and no internally consistent value could be used to replace it. These edits were performed before the edits listed in Section 8.2.4.1 were completed. For respondents who had changes to their rosters that were due to the edits described below, the changes to age and gender that were due to the edits described in Section 8.2.4.1 were checked to make sure that they did not impact the decision to implement the edits below. The relationship code in these instances was set to a bad data code.

1. More than one roster member aged 15 years or older was listed as the respondent's unmarried partner or as the respondent's spouse. This situation should have been covered by consistency check #1 listed in Section 8.2.1.2.
2. A roster member aged 15 years or older was identified as a spouse and another was identified as an unmarried partner. In this case, the spouse code was maintained and the unmarried partner code was set to bad data. This situation should have been covered by consistency check #1 listed in Section 8.2.1.2.
3. The roster member was the respondent's parent, but was younger than the respondent. This situation should have been covered by consistency check #2 listed in Section 8.2.1.2. This edit would have been automatic for respondents younger than 15 years old.
4. The roster member was the respondent's child, but was older than the respondent. This situation should have been covered by consistency check #3 listed in Section 8.2.1.2. This edit would have been automatic for respondents younger than 15.
5. The roster member was the respondent's biological parent, but was less than 13 years older than the respondent. This situation should have been covered by consistency check #9 listed in Section 8.2.1.2.
6. The roster member was the respondent's biological mother, but was more than 60 years older than the respondent.
7. The roster member was the respondent's biological child, but was less than 13 years younger than the respondent. This situation should have been covered by consistency check #10 listed in Section 8.2.1.2.
8. A respondent had a biological sibling older than a biological parent, where the biological parent was at least 13 years older than the respondent. If this situation occurred, the relationship code of the "sibling" was set to missing. If the age difference between the biological sibling and the respondent was more than 25 years, then a consistency check was triggered (consistency check #11 listed in Section 8.2.1.2).
9. A respondent had a biological parent younger than a biological sibling, where the biological parent was less than 13 years older than the respondent. If this situation occurred, the relationship code of the "parent" was set to missing. As with the previous edit, this edit was partially covered by consistency check #11 listed in Section 8.2.1.2.

10. The roster member was the respondent's child-in-law, but was at least 10 years older than the respondent. This situation should have been covered by consistency check #7 listed in Section 8.2.1.2.
11. The roster member was the respondent's parent-in-law, but was at least 10 years younger than the respondent. This situation should have been covered by consistency check #8 listed in Section 8.2.1.2.
12. The roster member was the respondent's parent-in-law or child-in-law, but either the roster member or the respondent was younger than 15 years old. This situation should have been covered by consistency check #6 listed in Section 8.2.1.2.
13. The respondent had two or more children-in-law, but had no children in the household. The in-law codes were all set to missing.
14. The roster member was the respondent's grandchild, but the respondent or respondent's spouse (if applicable) was 25 years old or younger. This situation should have been covered by consistency check #12 listed in Section 8.2.1.2.
15. The roster member was the respondent's grandchild, but the respondent's parents lived in the household. Also, the respondent had no children in the household and was less than 24 years older than the roster member. As with the previous edit, if the grandchild was in fact older than the respondent, this error should have been covered by consistency check #3 listed in Section 8.2.1.2.
16. The roster member was the respondent's sibling and the previous roster member[123] was a parent, but the roster member's age was within 4 years of the age of the parent. If the sibling was a half- or step-sibling, an additional requirement was that there was only one parent.
17. The roster member was the respondent's grandparent or grandchild, but the age difference between the respondent or the respondent's spouse (if applicable) and the roster member was less than 20 years. If the roster member was a "grandchild" who was older than the respondent, then this situation was covered by consistency check #3 listed in Section 8.2.1.2. Similarly, if the roster member was a "grandparent" who was younger than the respondent, then this situation was covered by consistency check #2 listed in Section 8.2.1.2. If the age difference was less than 30 years, this was covered by consistency check #12 in Section 8.2.1.2.
18. If the respondent had two parents, but both parents were listed as biological mothers or both parents were listed as biological fathers, the roster genders of both roster members were set to missing.

8.2.4.3 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Child Code)

In Section 8.2.4.2, illogical relationship codes were set to bad data. Often, this occurred because the interviewer used someone other than the respondent as the reference person for one or more roster members. In some of these cases, the structure of the roster could have been used

[123] A "previous roster member" is the member who immediately precedes the member of interest in the roster.

to determine the appropriate relationship code for that individual. Edits where the illogical code was "child" are listed below.

1. The interviewer might have put a roster member after the respondent's parent in the household roster. If the relationship code for that roster member was given as "child," the relationship code was illogical if the age made it impossible for the roster member to be the respondent's child (see #4 in Section 8.2.4.2). In fact, if more than one "child" was listed after the respondent's parent, each would be listed as illogical. However, it was likely that the interviewer was making the reference to the respondent's parent rather than the respondent. In this case, if the child relationship was not a stepchild and the age difference between the respondent's parent and the "child" was at least 12 years, then the relationship code was changed to sibling.
2. In some cases, the interviewer's entry for a roster member listed as "child" might simply be a typographical error, for example, where the "3" (child) should be a "2" (parent) (see Table 8.2 for all the relationship codes). Interviewers usually corrected such errors when a consistency check was triggered in cases where the child was older than the parent or the child was a biological child who was less than 12 years younger than the parent (see Section 8.2.4.2). However, in cases where the interviewer insisted on the code, or where the child was younger than the respondent, but was less than 12 years younger than the respondent and was not biological, these typographical errors were more difficult to detect. If the respondent was living with parent(s) and unmarried and not living with an unmarried partner, and the roster member was not 12 or more years younger than the respondent, then the relationship code was changed to sibling.
3. Both sides in a selected pair[124] were respondents aged 18 or younger, both sides identified parents in the household, and one side had an illogical child code. When the number of illogical child codes was added to the number of siblings on one side, the sum was equal to the number of siblings on the other side. If the age of the roster member was younger than 25 years, then the relationship code was changed to sibling.
4. A roster member was listed as the respondent's child who was not more than 12 years younger than the respondent and the respondent was 25 or younger. The previous roster member was listed as "grandparent." The "child" was in reference to the respondent's grandparent and was considered either the respondent's parent or the respondent's uncle or aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2), then no uncles or aunts lived in the household. If a pair was selected and no nonimmediate family codes were found in either pair member's roster, then in either of these cases the relationship code was set to parent. Otherwise, the relationship code was set to missing.

[124] A selected pair has two rosters where each respondent is from the same household. A "side" refers to one of the two rosters that make up a selected pair.

8.2.4.4 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Spouse Code)

The interviewer also could have used an incorrect reference person with illogical spouse codes. This error occurred most frequently when a selected child had a parent with a spouse (the other parent) or unmarried partner. Rather than identifying this individual as a "parent" or "other nonrelative," the interviewer identified the roster member as a spouse or unmarried partner of the child, even though the interviewer intended that the point of reference be the child's parent rather than the child. This manifestation of the illogical spouse code, along with others, is described below. Many of these edits were covered by consistency checks #4 and #5 listed in Section 8.2.1.2, provided either the respondent or the roster member was 16 or younger.

1. Both sides in a selected pair identified that they had a spouse or unmarried partner, but the two respondents were not part of a spouse-spouse pair. This legitimately could have occurred only if there were multiple spouse-spouse pairs in the household. In this edit, an attempt was made to identify cases with a single spouse-spouse pair in the household, where one pair member had a correctly identified spouse or unmarried partner and the other pair member had an incorrectly identified spouse or unmarried partner. If the younger respondent, who was 21 years old or younger and at least 10 years younger than the older respondent, indicated a parent, and the older respondent indicated neither parents nor parents-in-law, then the older respondent should be considered either the younger respondent's parent or the parent's spouse or unmarried partner. If the misidentified code was "spouse," then the code was changed to "parent." However, if the misidentified code was "unmarried partner," then the roster member may or may not be considered the parent of the respondent. In most cases where the misidentified unmarried partner was the respondent's parent's unmarried partner, the code was changed to parent. The exception occurred when (1) the unmarried partner of this respondent's parent was the other respondent selected in a pair, and (2) the unmarried partner did not indicate that the other pair member selected was his or her child in the parenting experiences question, FIPE3. In this instance, the relationship code was changed to a special code indicating that the roster member was an unmarried partner of the respondent's parent.
2. As in the previous edit, both sides in a selected pair identified a spouse or unmarried partner, but were not part of a spouse-spouse pair, and there was only a single spouse-spouse pair in the household. In this edit, both sides incorrectly identified the spouse or unmarried partner. In most cases, the pair was a sibling-sibling pair. If both respondents were younger than 21, both indicated a parent in the household, and the age difference between the respondents and their respective "spouse or unmarried partner" was unusually large, then on each side the misidentified spouse or unmarried partner should have been considered a spouse or unmarried partner of the respondent's parent. If both misidentified codes were "spouse," then both codes were changed to "parent." As stated in the previous edit, if both misidentified codes were "unmarried partner," then it was not clear whether each misidentified code should have been "parent." The rules used to determine whether the roster member was the respondent's parent were the same as in edit #1. The same special code as in the previous edit was used to identify an unmarried partner of the respondent's parent.

Hence, the incorrectly identified "spouse or unmarried partner" code was changed for each respondent in the pair to either "parent" or the aforementioned special code.

3. In this edit, only one side in a selected pair identified a spouse (not unmarried partner), but the spouse was identified even though (1) the respondent was younger than 15, (2) the spouse was younger than 15 and the other pair member did not have a spouse, or (3) the respondent was younger than 18 but responded that he or she was "never married" in the core part of the questionnaire, and the respondent did not have any parents-in-law in the household. If the respondent listed one parent, but the other pair member listed two parents, then the pair was a sibling-sibling pair and the relationship code was in reference to the parent. If the respondent listed one fewer sibling than the other pair member, then the pair was a sibling-sibling pair and the spouse code was a typographical error (meant to be a sibling, with a code of "4" instead of "5").

4. Only one side in a selected pair identified an unmarried partner, but the unmarried partner was identified even though (1) the respondent was younger than 15 or (2) the unmarried partner was younger than 15. If the respondent listed one parent, but the other pair member listed two parents, then the pair was a sibling-sibling pair and the relationship code was in reference to the parent's unmarried partner. In this case, the relationship code was changed to parent. If the respondent listed one fewer sibling than the other pair member and the age difference between the respondent and the roster member identified as the unmarried partner was less than 15 years, then the pair was a sibling-sibling pair and the unmarried partner code was changed to sibling.

5. Both sides in a pair identified the same household member as spouse or unmarried partner. If the previous roster member on one of the sides was a sibling, then the spouse or unmarried partner should be considered the sibling's spouse or unmarried partner. The spouse or unmarried partner relationship code was changed to bad data. If both sides had a previous roster member who was a sibling, then it was not clear to which pair member the spouse or unmarried partner belonged. To maintain proper counts, the spouse or unmarried partner code for the youngest pair member was changed.

6. A spouse or unmarried partner was identified even though (1) the respondent had one parent in the household who was the roster member listed before the spouse or unmarried partner; (2) the respondent either was younger than 17 years old or was between 17 and 20 years old and the spouse or unmarried partner was older than the respondent's parent; and (3) the respondent was more than 15 years younger than the spouse or unmarried partner. In the case of the misidentified spouse, the "spouse" of the respondent was considered the respondent's other parent. In the case of the misidentified unmarried partner, the "partner" of the respondent was considered the unmarried partner of the respondent's parent. The code was changed to "parent." For a household member with a spouse code who was aged 16 years or younger, this edit should have been covered by consistency check #4 listed in Section 8.2.1.2.

7. In cases where the respondent was younger than 15 years old, he or she identified a spouse or unmarried partner, and the above edits did not apply, the relationship code was set to bad data. In cases where the roster member was younger than 15, the roster

member was identified as a spouse or unmarried partner, and the above edits did not apply, the relationship code and roster member's age were set to bad data. This should have been covered by consistency checks #4 and #5 listed in Section 8.2.1.2.

8.2.4.5 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Sibling Codes)

If the relationship code indicated that one of the other roster members was the respondent's sibling, but the age difference between the sibling and the respondent was at least 20 years, then the sibling relationship code was suspicious. If the previous roster entry was either the respondent's child or another sibling with the same characteristics, and either the respondent did not have parents in the household or the parent was a mother and the age difference between the mother and the sibling was more than 50 years, then the sibling relationship codes were referencing the respondent's children's relationships to each other. The relationship codes were therefore changed to "child." Age differences greater than 25 years among biological siblings would have been covered by consistency check #11 in Section 8.2.1.2.

8.2.4.6 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical Grandchild Codes)

If the relationship code indicated that one of the other roster members was the respondent's grandchild, but the respondent was too young to have a grandchild (25 or younger), it was possible that the roster member was a grandchild of a previous roster member. If two young respondents were selected where both identified the same grandparents and the same parents, and the respondent on the other side had siblings, then the grandchild should be considered the respondent's sibling. If this was not established, then the roster member could be the respondent's sibling or the respondent's cousin, and the code was set to bad data. If the grandchild was older than the respondent, then this edit would have been covered by consistency check #3 listed in Section 8.2.1.2. If the age difference between the grandchild and the respondent was less than 30 years, then this edit would have been covered by consistency check #12 listed in Section 8.2.1.2.

8.2.4.7 Edits to Relationship Codes: Changes to Different Values (Incorrect Reference Person: Illogical In-Law Codes)

In some situations, the in-law code was incorrectly used because the respondent was not using himself or herself as the reference person. In such cases, either the child-in-law was the child of someone else in the roster other than the respondent or the respondent was referring to himself or herself as the parent-in-law of the roster member. An in-law code was deemed incorrect if a roster member was listed as the respondent's child-in-law who was not more than 12 years younger than the respondent and the respondent was 25 or younger. If the relationship code was listed as child-in-law, and the previous roster member was listed as grandparent, then the child-in-law was in reference to the respondent's grandparent and should have been considered either the respondent's parent or the respondent's uncle or aunt. If the roster member's age was at least 12 years older than the respondent and there were no nonimmediate family codes (7, 12, 13, or 14 as described in Table 8.2), then no uncles or aunts lived in the household. If a pair was selected, no nonimmediate family codes were found in either pair member's roster. In

either of these cases, the relationship code was set to parent. Otherwise, no certainty was associated with the relationship code, and this code was set to missing.

8.2.5 Final Edited Respondent-Level Detailed Roster Variables

The unedited roster variables contained information for each roster member: age, gender, relationship to respondent, and a 0/1 variable that indicated whether the roster member was the other member selected in a pair (Table 8.1 provides an example). This information could be captured for up to 25 members of a household. Within the CAI instrument, separate variables were created to collect this information for male and female household members and for household members with ages reported in years as opposed to months. When the edited versions of these variables were created, this information was combined for each household member into four variables, one for each attribute (i.e., age, gender, relationship to respondent, and pair status). The edits listed in Section 8.2 were incorporated into the values of the detailed roster variables, called ROSAGE1-ROSAGE25 (roster age), ROSSEX1-ROSSEX25 (roster gender), ROSRLT1-ROSRLT25 (relationship to respondent), and ROSMSL1-ROSMSL25 (0/1 indicator: other member selected, pair members only). Additional variables were also created: PRNTYP1-PRNTYP25 (type of parent: biological, adoptive, etc.), SIBTYP1-SIBTYP25 (type of sibling: biological, adoptive, etc.), CHDTYP1-CHDTYP25 (type of child: biological, adoptive, etc.), and TWNTYP1-TWNTYP25 (type of twin: identical, fraternal, or neither).

Final edited versions of the respondent-level detailed roster variables were used to derive (or, at a minimum, to calculate bounds when data were missing) the household composition variables described in Section 8.3.

8.3 Editing and Imputation Procedures for Roster-Derived Household Composition Variables

8.3.1 Creation of Edited Roster-Derived Household Composition Variables

This section discusses the creation of edited versions of the roster-derived household composition variables. After replacing apparently erroneous information in the roster with missing values, the number of individuals with various characteristics in each roster was determined. These counts were recorded in the edited roster-derived household variables shown in Table 8.3. If any information in the roster was missing, the roster-derived variable was set to missing. However, if some of the roster records for a respondent's household had missing data, then roster records with nonmissing data for that household were used to limit the possible values to which the missing roster-derived variable could have been imputed. Details on the imputation of the roster-derived household variables are provided in Section 8.3.2. If two respondents were selected in a single household as part of a pair, then the information from one pair member was not used to edit that of the other pair member.

The respondent's household size was assumed to equal the total number of rostered persons in the household, TOTPEOP, as shown in Table 8.3. The value of TOTPEOP was expected to equal the value of QD54 in most cases. However, in some cases, the original self was misidentified and no other roster members were close to matching the respondent's age and gender. In these cases, an extra roster member was added to correspond to the respondent (the

self) so that the value of TOTPEOP was 1 greater than the value of QD54. For other cases, the respondent did not enter a value for QD54, and thus TOTPEOP and all the roster-derived variables were missing. Finally, it was possible that duplicate entries were put into the household roster so that the value of TOTPEOP would be determined by excluding the duplicates from the roster. This latter situation was usually impossible to detect, unless the respondent had two biological fathers or two biological mothers of exactly the same age. In this instance, the extra biological parent of the same gender was dropped from the roster, and the value of TOTPEOP was reduced to 1 less than the value of QD54.

The variables KID17 (number of persons in the household aged 17 or younger) and HH65 (number of persons in the household aged 65 or older) were simple counts based on the roster ages and did not account for the relationships of the individuals to the respondent. If some of the roster members had missing ages, the values of KID17 and HH65 also were missing, regardless of whether some of the roster members were eligible to be part of the count. In these instances, the imputed values for KID17 and HH65 were restricted based on the nonmissing information available in the roster, as explained in Section 8.3.2.2. However, if the roster member was missing a relationship code, but not an age, then that roster member was still eligible to be included in these variables.

The variable FAMSKIP was an indicator of whether the respondent's household contained any other family members. It was created based on the relationship codes of the roster members. If one or more of the roster members had a missing relationship code, and no other family members were in the respondent's household, then the value of FAMSKIP was set to missing. However, if one of the nonmissing roster member's relationship codes indicated that the household contained one of the respondent's family members, then the value of FAMSKIP was not missing, even if other roster members had missing relationship codes.

The variables FMLYSIZE (number of respondent's family members in the household, including foster relationships), FAMSIZE (number of respondent's family members in the household, excluding foster relationships), KIDFMLY (number of respondent's family members in the household aged 17 or younger, including foster relationships), and KIDFAMSZ (number of respondent's family members in the household aged 17 or younger, excluding foster relationships) were simple counts based on the relationships of the individuals to the respondent and the ages in the respondent's household roster. FMLYSIZE and KIDFMLY were created to determine appropriate measures of poverty levels, using Federal poverty definitions starting in 2006. FAMSIZE and KIDFAMSZ were used in the 2003-2005 surveys. The definition of "family" for FAMSIZE and KIDFAMSZ was a little different from that used for other roster variables; foster relationships were not considered family relationships. If some of the roster members had missing ages or missing relationship codes, the values of FMLYSIZE, FAMSIZE, KIDFMLY, and KIDFAMSZ were set to missing, even though some of the roster members might have been eligible to be part of the count. In these instances, the imputed values were restricted based on the nonmissing information available in the roster, as explained in Section 8.3.2.2.

Eleven other roster-derived variables were created that used both the age and relationship codes of the roster members. All of the roster-derived variables and their definitions are summarized in Table 8.3. Except for FAMSKIP, each of these variables was missing if the age

or relationship codes for at least one roster member in a respondent's household were missing. FAMSKIP could be coded despite missing values if there was at least one nonmissing family relationship code in the roster. Edited versions of the roster-derived variables were also used in the editing procedures applied to the creation of the edited proxy variables, described in Section 8.4.

8.3.2 Imputation for Roster-Derived Household Composition Variables

Of the three groups of edited roster variables described in the introduction to this chapter, the only group that underwent any imputation at all was a subset of the roster-derived household composition variables. Each of the eight variables in this subset formed its own single-member imputation set[125] and tended to have item response rates of more than 99 percent. Table A.28 in Appendix A has details on the rates of missingness for the variables that were imputed. The single RP/single PRD type of PMN, described in Section 3.4.1, was used to impute nonmissing values among these eight variables in the order shown in Table 8.3. The order was important, as imputation-revised variables from earlier in the sequence were frequently used to assist with imputation-revised variables later in the sequence.

Section 8.3.2.1 describes the imputation process applied to the first edited variable, TOTPEOP. Section 8.3.2.2 summarizes the imputation processes applied to the other seven variables. Since the processes applied to the other seven variables are very similar to the process applied to TOTPEOP, Section 8.3.2.2 will only list divergences from the process that was applied to the TOTPEOP variable.

8.3.2.1 Imputation for TOTPEOP (Imputation Set 1)

The first imputation set included a single variable, TOTPEOP. The analogous imputation-revised variable IRHHSIZE was created using the single RP/single PRD type. There were no noteworthy deviations from this general approach. Section 8.3.2.1.1 describes the RP step, Section 8.3.2.1.2 describes the PRD step, and Section 8.3.2.1.3 describes the hot-deck step. As is true for all the roster-derived household composition variables that underwent imputation, the item response rate was very high (more than 99 percent).

8.3.2.1.1 Response Propensity Step

The response propensity model for imputation set 1 utilized the preliminary analysis weight, PANALWT. All respondents were in the domain for the TOTPEOP variable. A domain member was considered an item respondent if and only if TOTPEOP was nonmissing. See Tables D.49, D.50, D.51, and D.52 in Appendix D for details of the covariates used in the RP models for this variable.

8.3.2.1.2 Prediction Step

TOTPEOP was a count variable. It was assumed to have a Poisson distribution, and the parameters for the models were estimated using the adjusted weights that are outputs of the RP

[125] An imputation set is a set of variables for which a single donor is used in the final hot-deck step. Chapter 3 describes this concept more fully.

step and using Poisson regression as implemented by the LOGLINK procedure in SUDAAN software.[126] The single predicted mean used in the subsequent hot-deck step was the predicted number of people in the household.

8.3.2.1.3 Hot-Deck Step

The hot-deck step for the TOTPEOP variable was the simplest one used in the NSDUH. There were no logical constraints, and the only likeness constraint was the delta constraint. The predictive mean vector was actually a scalar. Every item nonrespondent was handled on the first attempt to find a donor. Additional details on the hot-deck step for TOTPEOP are available in Appendix E.

8.3.2.2 Imputation for Other Roster-Derived Household Composition Variables That Underwent Imputation (Imputation Sets 2 through 8)

Like TOTPEOP, the remaining seven roster-derived household composition variables that underwent imputation (from Table 8.3: KID17, HH65, FAMSKIP, FMLYSIZE, KIDFMLY, FAMSIZE, and KIDFAMSZ) were handled separately using the single model type of PMN and utilized the preliminary analysis weight, PANALWT, in the RP step. The methods were very similar, with only a few exceptions, as follows:

- FAMSKIP was a dichotomous variable, not a count variable. Therefore, its PRD model was a logistic regression model as implemented by the RLOGIST procedure in SUDAAN. The single predicted mean used in the later hot-deck step was the predicted probability that the respondent did not have any other family members in his or her household.
- Bounds were determined for every other variable except FAMSKIP. These bounds were based both on nonmissing information in the roster and on previously imputed variables. For each of these variables, a single logical constraint was used in the hot-deck step, which required the donor to have a value within the bounds.
- Previously imputed roster-derived household composition variables were frequently used in likeness constraints in the hot-deck steps.

Tables D.49, D.50, D.51, and D.52 in Appendix D provide details of the covariates used in the RP models for these variables.

8.4 Editing the Proxy Variables

This section describes the creation of edited proxy variables, as listed in Table 8.4. Section 8.4.1 describes the creation of an indicator variable, EDFAM18, which was used to determine skip patterns and missing codes for the five edited proxy variables. Sections 8.4.2 and 8.4.3 describe the editing processes for each value of EDFAM18.

[126] Details about the LOGLINK procedure and additional references are provided in the *SUDAAN Language Manual, Release 10.0* (RTI International, 2008).

All survey respondents were allowed to choose someone from the household to be their proxy as long as the following conditions were met:

1. There was more than one person in the household.
2. The eligible person was a relative (not a boarder, roommate, or some other nonrelative).
3. The eligible person was aged 18 or older.

Table 8.4 shows the correspondence between the five questionnaire items in the proxy section of the questionnaire and the corresponding edited variables. Except for QP02 and its edited variable PRXRELAT, the valid questionnaire responses were "1 = Yes" and "2 = No." QP02 and PRXRELAT had multiple responses ranging from 1 to 21, with each level representing the relationship of the proxy to the respondent. The levels of PRXRELAT are shown in Table 8.5.

Table 8.5 Levels of PRXRELAT

PRXRELAT	Relationship of Proxy Member	Gender of Proxy Member
1 = Father	Parent	Male
2 = Mother	Parent	Female
3 = Son	Child	Male
4 = Daughter	Child	Female
5 = Brother	Sibling	Male
6 = Sister	Sibling	Female
7 = Husband	Spouse	Male
8 = Wife	Spouse	Female
9 = Male Unmarried Partner	Unmarried partner	Male
10 = Female Unmarried Partner	Unmarried partner	Female
11 = Son-in-law	Child-in-law	Male
12 = Daughter-in-law	Child-in-law	Female
13 = Grandson	Grandchild	Male
14 = Granddaughter	Grandchild	Female
15 = Father-in-law	Parent-in-law	Male
16 = Mother-in-law	Parent-in-law	Female
17 = Grandfather	Grandparent	Male
18 = Grandmother	Grandparent	Female
19 = Other Male Relative	Other relative	Male
20 = Other Female Relative	Other relative	Female
21 = Other Adult Relative	Other relative	Male or Female

8.4.1 Edited Indicator of Potential Proxies in Household (EDFAM18)

As described in Section 8.3.1, a binary variable (FAMSKIP) was created that indicated whether the respondent's household roster included other family members. If the presence or

absence of other family members was unknown because of a missing household size or missing values in the roster, FAMSKIP could not be determined. A similar variable was created to identify households where the respondent's household roster included other family members aged 18 years or older ("adult" family members), any one of whom could potentially serve as a proxy for the respondent. The edited indicator was called EDFAM18, where "1" indicated that no potential proxy existed in the respondent's household, "0" indicated otherwise, and "98" indicated unknown.

8.4.2 Editing the Proxy Variables when EDFAM18 = 1

In most cases, a value of EDFAM18 = 1 implied that the respondent was skipped out of the proxy questions because no potential proxy existed in the household. In these cases, all of the proxy variables were given a legitimate skip code (99). Two situations could occur, however, where adult family members were incorrectly identified in the household roster: (1) the respondent had not identified any adult family members in the household but had nonfamily members in the household whose ages were not known; and (2) the unedited household roster indicated that one potential proxy existed in the household but editing changed the age of this single potential proxy to younger than 18. In these cases, the respondent was allowed to answer the proxy questions even though the value of EDFAM18 was 1 (i.e., the final edited household roster indicated that no potential proxy existed in his or her household). Moreover, in these situations, the interviewer indicated that none of these household members who were incorrectly identified as adult family members were proxies. However, the "no" value in the first unedited proxy variable (QP01) was replaced by a logically assigned legitimate skip (89) in the corresponding edited variable (PRXABLE2). For cases where PRXABLE2 was set to 89, all of the edited proxy variables corresponding to the unedited proxy variables, which followed QP01, were given legitimate skip codes (99). These were cases in which the respondent answered the proxy module (questions about a proxy) but the interviewer indicated that they were not proxies, so no proxy should have actually answered the health insurance and income modules.

8.4.3 Editing the Proxy Variables when EDFAM18 = 0

If EDFAM18 was 0, the proxy variables were edited as follows:

1. If the unedited proxy variables had legitimate nonmissing values (i.e., not replaced by a logically assigned legitimate skip), the edited proxy variables (except PRXRELAT) were set to those nonmissing values.
2. If any of the unedited proxy variables (except PRXRELAT) had a value of 2 ("no"), then all of the variables that followed were edited to legitimate skips.
3. If any of the unedited proxy variables had a value of "don't know" or "refused," then the corresponding edited variable and all the edited variables that followed were given a "don't know" or "refused" code (94 or 97).
4. If any of the unedited proxy variables did not have a value and a legitimate skip code could not be applied, then the corresponding edited variable and all the variables that followed were given a "no answer" code (98).

8.4.4 Additional Editing for PRXRELAT

In addition to these, more detailed rules were used to assign values to PRXRELAT. The value of QP02, which identified the proxy for the respondent, was chosen directly from the respondent's household roster. A list of adult family members (a proxy roster) was shown to the respondent, and the respondent was asked to select the family member who could best answer the health insurance and income modules. In the cases where the proxy roster included a large number, only the first nine adult family members were listed. Once the proxy roster was established, the number selected in QP02 was matched to the corresponding person in the proxy roster.

9. Editing and Imputation for the NSDUH Income Variables

9.1 Introduction

This chapter describes procedures for editing and imputing the income variables in the 2012 National Survey on Drug Use and Health (NSDUH) interview. Unlike other sections of the interview that asked about behaviors or situations in the 12-month period prior to the interview, the reference period for the income module on the 2012 NSDUH was the previous calendar year (i.e., January 1, 2011, through December 31, 2011). Respondents (or other household members serving as proxies[127]) were asked whether they (or other family members living in the dwelling unit [DU], if applicable) received income or benefits from specific sources, such as social security or wages earned at a job or business, during the previous calendar year. These questions are subsequently referred to as "source-of-income" questions in this chapter. In addition, the income section of the interview included questions about the dollar amounts of total personal income and total family income (if applicable) during the previous calendar year.

As had been the case in the 2008-2011 surveys, separate questions to ascertain personal-level and other-family-level responses for binary (i.e., yes/no) source-of-income variables were not asked in the 2012 survey, nor were there separate questions about income from child support, interest/investment income, and other income. However, since 2008, respondents have been asked questions about binary (i.e., $20,000 or more or less than $20,000) and finer categories of total annual income at both the personal and family levels. A comparison between the 1999-2007 and 2008-2012 sets of income questions is shown in Table 9.1. See Section 3.4 of the 2008 imputation report (Ault et al., 2010) for a more detailed explanation of the changes to the income questions over the years.

Consistent with most of the imputation-revised variables that were discussed in the previous chapters of this report, imputations for the income variables in the 2012 NSDUH were performed using the predictive mean neighborhood (PMN) methodology detailed in Chapter 3. The edits that were applied to the income variables prior to imputation are described in Section 9.2.

Table 9.1 Comparison between 1999-2007 and 2008-2012 Sets of Income Questions

Income Questions Included in NSDUH	1999-2007 Surveys*			2008-2012 Surveys	
	Personal Level	Other Family Member Level	Family Level	Personal Level	Family Level
Social Security	Yes	Yes	No	No	Yes
Supplemental Security Income	Yes	Yes	No	No	Yes

[127] Reference is made only to "respondents" in the remainder of this chapter. However, readers are advised that the income information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the income questions for the respondent. See Chapter 8 for more information about proxy variables.

Table 9.1 Comparison between 1999-2007 and 2008-2012 Sets of Income Questions (continued)

Income Questions Included in NSDUH	1999-2007 Surveys*			2008-2012 Surveys	
	Personal Level	Other Family Member Level	Family Level	Personal Level	Family Level
Welfare Payments	Yes	Yes	No	No	Yes
Other Welfare Services	Yes	Yes	No	No	Yes
Investment Income	Yes	Yes	No	No	No
Child Support Payments	Yes	Yes	No	No	No
Wages	Yes	Yes	No	No	Yes
Other Income	Yes	Yes	No	No	No
Food Stamps	No	No	Yes	No	Yes
Months on Welfare	No	No	Yes	No	Yes
Binary Total Income	Yes	No	Yes	Yes	Yes
Finer Category Total Income	Yes	No	Yes	Yes	Yes

* In the 2006-2007 surveys, 5 percent of the selected sample was asked a reduced set of income questions, which are identical to the set of questions asked in the 2008-2012 surveys. For detailed explanations of these subsamples, consult the 2006 and 2007 Methodological Resource Books.

9.2 Editing for Binary and Finer Category Income Variables

9.2.1 Binary Income Variables

As stated previously, the source-of-income variables were binary (i.e., yes/no). Information for most of these variables was captured through single questions about a given source of income. The exception to this single-question format was for measurement of the number of months on welfare (questions QI12AN and QI12BN and the edited variable WELMOS). Question QI12AN was asked if respondents did not report any family members' receipt of food stamps in question QI07N; otherwise, question QI12BN was asked. In asking about the number of months on welfare, question QI12BN included the additional phrase, "not including food stamps."

Creation of WELMOS also was dependent on respondents' answers in the questions corresponding to the relevant edited variables for family members' receipt of cash assistance from a State or county welfare program (FAMPMT) and receipt of other (noncash) welfare or public assistance (FAMSVC). If FAMPMT or FAMSVC indicated that respondents received welfare payments, WELMOS was assigned the values from questions QI12AN and QI12BN. Otherwise, if family members in the household (including the respondent) definitely did not receive welfare payments or noncash benefits (e.g., job training, help with child care), WELMOS was assigned a legitimate skip code.

In addition to the source-of-income variables, binary variables were created from a pair of questions that asked whether the respondent's personal total income or the respondent's total family income was $20,000 or more. For this pair of questions (QI20N and QI22, corresponding to edited variables PINC1 and FINC1), the second question in the pair applied to the entire family. For answering these questions about total personal or family income, respondents were shown a list of other sources of income that they had not been previously asked about (e.g., child

support payments). Respondents were instructed in answering these questions about total income to consider both of these other sources of income and the sources of income they had been asked about previously.

If a respondent had no other family members living in his or her DU, FINC1 was assigned a legitimate skip code (Section 2.4.2). QI22 also was skipped and FINC1 was assigned a legitimate skip code if respondents reported in QI20N that they had personal incomes of $20,000 or more. A third binary total family income variable, FAMINC1, was created and was equal to either PINC1 or FINC1, depending on whether other family members were present in the household.

9.2.2 Finer Category Income Variables

Respondents also were asked to identify, both for themselves and for their families, income within the two general categories of "$20,000 or more" or "Less than $20,000." If respondents answered binary total income questions as "Less than $20,000," they were asked to report a finer category of income from incomes less than $1,000 (including a net loss of income) up to $20,000 by increments of $1,000. Similarly, respondents who answered binary total income questions as "$20,000 or more" were asked to report a finer category of income from $20,000 up to $50,000 in increments of $5,000. If respondents' or their families' incomes were greater than $50,000, they could select the following additional categories: "$50,000 to $74,999," "$75,000 to $99,999," or "$100,000 or more." (Prior to 2004, the maximum income category was "$75,000 or more.")

The variable PINC2 for finer categories of personal income was created from responses to questions QI21A (for personal incomes of less than $20,000) and QI21B (for personal incomes of $20,000 or more). The variable FINC2 for finer categories of family income was created from questions QI23A (for family incomes of less than $20,000) and QI23B (for family incomes of $20,000 or more). Income categories in PINC2 and FINC2 ranged from 1 (less than $1,000 [including loss]) to 29 ($100,000 or more).

Questions QI23A or QI23B were skipped if the respondent had no other family members in the household. Therefore, when the imputed household roster variable IRFAMSKP indicated that no other family members were living in the household, FINC2 was assigned a legitimate skip code of 99; that included situations in which data existed in questions QI23A or QI23B but IRFAMSKP subsequently indicated that no other family members lived in the household.

A third total family income variable with finer categories of income, FAMINC2, also was created. If other family members were living in the household, FAMINC2 was set to be equal to the response from the mutually exclusive pair of total family income questions QI23A and QI23B. If no other family members lived in the household, FAMINC2 was set to be equal to the responses from questions QI21A or QI21B. Thus, unlike FINC2, FAMINC2 was not assigned legitimate skip codes if no other family members lived in the household. Since 2010, respondents who reported in question QI21B that their total personal income was $100,000 or more were not asked question QI23B about their total family income. FAMINC2 was assigned a code of 29 ($100,000 or more) in this situation.

9.2.3 Additional Edits to Income Variables for Inconsistencies or Logical Inferences

Relatively little additional editing was done to the income variables, aside from assigning legitimate skip codes (see Section 2.4.2). This section describes the additional edits that were implemented because of inconsistencies in the data or to make logical inferences.

As noted in Section 4.2.3, interviewers entered information at the beginning of the interview about the State where the DU was located. This information was used in the income module to fill in the State-specific names for Temporary Assistance for Needy Families (TANF) programs in question QI08N. If the edited variable STATELOC for the State where the DU was located had been set to bad data because of interviewer errors, any nonblank values in the variable corresponding to question QI08N about cash assistance from a State or county welfare program (FAMPMT) were set to bad data. This edit was implemented because the incorrect State residence information could result in respondents not being asked about cash assistance from the TANF program in the States where they actually lived. In turn, if FAMPMT had been set to bad data because STATELOC had been set to bad data, and FAMSVC did not indicate welfare assistance, then nonblank values in WELMOS also were set to bad data.

Two types of edits were implemented for overall family income variables based on data about personal income:

- If the binary family income variable FINC1 indicated that the family income was less than $20,000, and the corresponding personal income variable PINC1 had a missing value, then the respondent's personal income (PINC1) was logically inferred to be less than $20,000. Similarly, if the binary personal income variable PINC1 was $20,000 or more, and the family income variable FINC1 was skipped, then the total family income variable FAMINC1 indicated that the family income was $20,000 or more.
- If the finer category for total personal income (PINC2) was greater than the corresponding income category in FINC2, then FINC2 and FAMINC2 were set to bad data. For example, if PINC2 indicated that the respondent had an income of $25,000 to $29,999, and the respondent reported a total family income of only $20,000 to $24,999, then FINC2 and FAMINC2 were set to bad data.

9.3 Imputation for the Income Variables

The imputation of income was separated into two imputation sets. The first set involved the imputation of all the binary income variables (i.e., "yes/no" questions about the following sources of income: social security, supplemental security income, welfare cash assistance, welfare noncash assistance, wages, and food stamps), the number of months on welfare (the only variable that was not binary in this imputation set), and a yes/no question regarding whether the respondent's income or the respondent's household family income was $20,000 or more (including income from the sources referenced in the previous questions). This first set was processed using the single response propensity (RP)/multiple prediction (PRD) PMN type described in Section 3.4.3. The second imputation set for finer categories of income consisted of imputing more specific income categories for the respondent and the respondent's family in the

household. This set was processed using the single RP/single PRD PMN type as described in Section 3.4.1.

The income variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation. See Table A.26 in Appendix A for details on the rates of missingness for these variables.

Imputations for all income variables were conducted separately within four age groups: 12 to 17, 18 to 25, 26 to 64, and 65 or older. The segregation into age groups was done to exploit the correlation between the income variables and age and to allow parallel processing of the data (thus reducing the time it takes to implement the procedures).

All income RP and PRD models used a State-rank variable as a covariate. States were ranked by the weighted proportion of respondents whose personal incomes during the prior calendar year were greater than or equal to $20,000. See Section 3.5 for a general discussion of State-rank variables.

9.3.1 Imputation for Binary Income Variables (Imputation Set 1)

The PMN imputation type used for the binary income variables was single RP/multiple PRD. The RP model is described in Section 9.3.1.1. The PRD model for the first binary income variable imputed each year, family social security (FAMSOC), is described in Section 9.3.1.2, and the provisional hot-deck step for FAMSOC is described in Section 9.3.1.3. The remaining PRD and provisional hot-deck steps described in Section 9.3.1.4 list deviations from the analogous steps for family social security. The final hot-deck step applied to all binary income variables is described in Section 9.3.1.5. Finally, a recode for the GOVTPROG variable, made from four imputation-revised binary income variables, is described in Section 9.3.1.6.

For the binary income models that predict whether a respondent had a given source of income, other sources of income were useful covariates. Therefore, provisionally imputed values were used as covariates in subsequent models within the set. The order in which missing values for the binary income variables were provisionally imputed is listed in Table 9.2.

Table 9.2 Imputation Order for Binary Income Variables (Imputation Set 1)

Income Type	Edited Variable Name(s)	Imputation-Revised Variable Name(s)
Family Social Security	FAMSOC	IRFAMSOC
Family Supplemental Security Income	FAMSSI	IRFAMSSI
Family Welfare Payments	FAMPMT	IRFAMPMT
Family Other Welfare Services	FAMSVC	IRFAMSVC
Family Wages	FAMWAG	IRFAMWAG
Family Food Stamps	FSTAMP	IRFSTAMP
Family Months on Welfare	WELMOS	IRWELMOS
Total Personal/Family Income	PINC1, FINC1, FAMINC1	IRPINC1, IRFINC1, IRFAMIN1

9.3.1.1 Response Propensity Step

The response propensity models for the binary income variables utilized the preliminary analysis weight, PANALWT. All respondents were in the domain for the binary income variables (i.e., eligible to have a valid value for these variables). For the single RP model, a domain member was considered an item respondent if and only if all of the variables listed in Table 9.2 were nonmissing. See Table D.53 in Appendix D for details of the covariates used in the RP models for these variables.

9.3.1.2 First Prediction Step (FAMSOC)

FAMSOC is a binary variable; after imputation, its only values are "yes" and "no" since all other missing data codes are replaced with imputed values. Using the adjusted weights that are outputs of the RP step, it was modeled using logistic regression and, in particular, the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean used in later hot-deck steps was the predicted probability that the respondent's family-in-household received income from social security in the preceding calendar year.

9.3.1.3 First Provisional Hot-Deck Step (FAMSOC)

The provisional hot-deck applied to the FAMSOC variable is a simplified version of the final hot-deck step for the binary income variables, because its only purpose is to fill in missing values so that FAMSOC can be used as a covariate in the PRD models for binary income variables imputed later in the sequence. Section 3.4.2 describes the concept of a provisional hot-deck in more detail. The final hot-deck step is described in Section 9.3.1.5 and in tabular form in Tables E.98, E.99, and E.100 in Appendix E. This provisional hot-deck step included the following:

- The predictive mean vector included only one element: the predicted mean from the preceding step.
- No logical constraints were used.
- The following likeness constraints were used:
  - IRFAMSKP of donor = IRFAMSKP of recipient. This is a likeness constraint in this hot-deck step, not a logical constraint. The creation of IRFAMSKP is described in Section 8.3.2 (LogC1 in Table E.98).
  - Donor's predicted mean must be within 5 percent of recipient's predicted mean (delta constraint; LikC3 in Table E.99).
  - Donor must match recipient with respect to whether there are adults aged 65 or older in the household. The variable used in this constraint was IRHH65, whose creation is described in Section 8.3.2 (LikC5 in Table E.99).
- In the first attempt to find a donor, all three likeness constraints were applied. In the second attempt, the delta constraint (#2 above) was removed.

The provisionally imputed version of FAMSOC was called INTFAMSOC. INTFAMSOC was used as a covariate in the rest of the binary income PRD models.

9.3.1.4 Analogous Prediction and Provisional Hot-Deck Steps for Remaining Binary Income Variables (FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, WELMOS, and FAMINC1)

PRD models were fit for FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, and FAMINC1[128] in the same manner as for FAMSOC, as described previously. Only the PRD model for WELMOS was different. The domain of the WELMOS model included only respondents who reported that their family-in-household received welfare payments and/or other welfare services during the preceding calendar year, as defined by FAMPMT and FAMSVC. Least squares regression (not logistic regression) was used, where the dependent variable was a standard logit,[129] such that $Y = \text{logit}(p)$ and p = number of months on welfare divided by 12. The REGRESS procedure in SAS-callable SUDAAN was used to fit the model. The predicted mean from the WELMOS model was the predicted probability of receiving welfare in a given month in the previous calendar year, given that the respondent received welfare payments and/or welfare services in the previous calendar year.

The provisional imputation steps for the other variables were implemented in the same manner as for FAMSOC, as described previously, except that the likeness constraints sometimes differed slightly. The following deviations are noted:

- Only the provisional hot-deck step for FAMSOC included the constraint involving the IRHH65 variable.
- For FAMPMT and FAMSVC, LikC6 in Table E.99 was used: donor must match recipient with respect to whether there are children younger than 18 in the household. The variable used in this constraint was IRKID17, whose creation is described in Section 8.3.2. This likeness constraint was used in both the first and second attempts to find a donor.
- For FAMWAG, LikC7 in Table E.99 was used: donor must match recipient with respect to whether there are adults aged 18 to 64 in the household. The variables used in this constraint were IRHHSIZE, IRHH65, and IRKID17. The creation of these variables is described in Section 8.3.2. This likeness constraint was used in both the first and second attempts to find a donor.
- Also for FAMWAG, donor must match recipient with respect to whether he or she was employed. The variable used in this constraint was EMPSTATY, whose creation is described in Section 5.3.1. If the recipient was employed full time or part time, then the donor must also have been employed full time or part time. If the donor was not employed full time or part time, then the donor must also not have been employed full time or part time. This likeness constraint was used in both the first and second attempts to find a donor.
- FAMINC1 did not undergo a provisional imputation step since it was the last variable in the set.

[128] The response variable for the model for total income was total family income (not total personal income), and the provisionally imputed version of WELMOS was not included as a covariate in the model.

[129] The Cox empirical logit was used when a person was on welfare for all 12 months.

9.3.1.5 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the binary income variables' final hot-deck step are available in Appendix E. This section explains the more general ideas behind the hot-deck step applied to these variables.

Because 10 imputation-revised variables are created in this step, and almost any subset of them can be missing, there are 639 missingness patterns for the binary income variables. The first 255 cover patterns where WELMOS is nonmissing; these are summarized as pattern 5 in Table E.100. The remaining 384 cover patterns where WELMOS is missing; these are summarized as patterns 1 through 4 in Table E.100.

- When WELMOS is nonmissing, any subset of FAMSOC, FAMSSI, FAMPMT, FAMSVC, FAMWAG, FSTAMP, PINC1, and FINC1 may be missing. However, if *none* of them are missing, then nothing is missing and no imputation is required. This leads to $2^8 - 1 = 255$ missingness patterns.[130]
- When WELMOS is missing, the elements of the predictive mean vector related to FAMPMT, FAMSVC, and WELMOS depend on the values of FAMPMT and FAMSVC. There are six relevant combinations of FAMPMT and FAMSVC (Table 9.3). For each of the six, any subset of FAMSOC, FAMSSI, FAMWAG, FSTAMP, PINC1, and FINC1 may be missing. This leads to $6 \times 2^6 = 384$ missingness patterns.
  - For the combinations labeled 1 in Table 9.3, the recipient is known to have received welfare payments and/or welfare services and neither is missing. Here, no predictive mean vector elements are required for FAMPMT and FAMSVC,[131] and there is no need to manipulate the element for WELMOS, which is already conditional on receipt of welfare as defined by FAMPMT and FAMSVC.
  - For combinations 2 and 3, the recipient is also known to have received welfare payments or welfare services, but one is missing. Here, a predictive mean vector element (i.e., a predicted mean) is required for the missing one of the two, but again there is no need to manipulate the corresponding element for WELMOS.
  - For combination 4, the recipient is known not to have received welfare payments, but welfare services is missing. Here, a predictive mean vector element is required for FAMSVC. The WELMOS predicted mean must be made conditional on the receipt of welfare services in the previous calendar year. Combination 5 is similar, but FAMPMT is the missing one.
  - For combination 6, both FAMPMT and FAMSVC are missing. Here, predictive mean vector elements are required for both. The WELMOS predicted mean must be made conditional on the receipt of welfare payments and/or welfare services. This probability is unknown, but can be crudely approximated by assuming the two are independent. Under that assumption, the probability of receiving either or

[130] Technically, if FAMPMT and FAMSVC were both missing, then WELMOS was necessarily missing. Therefore, $2^6 = 64$ out of these 255 missingness patterns are logically impossible and need not be set up.

[131] In general, in either the single RP/multiple PRD or multiple RP/multiple PRD types of PMN, only the elements of the predictive mean vector corresponding to missing responses were used (see Chapter 3).

both is one minus the probability of receiving neither, as expressed by (1 – PMT) × (1 – SVC).[132]

- If FAMPMT = "No" and FAMSVC = "No," then WELMOS is assigned a skip code and is therefore nonmissing. This combination is irrelevant for the purposes of this discussion, which only covers cases where WELMOS is missing.
- Combinations 1 through 3 in Table 9.3 correspond to pattern 1 in Table E.100, combination 4 corresponds to pattern 2, combination 5 corresponds to pattern 3, and combination 6 corresponds to pattern 4.

Table 9.3 Cross-Classification of FAMPMT and FAMSVC; Relevant Combinations for Binary Income Predictive Mean Vector when WELMOS Is Missing

		FAMSVC		
		Yes	**No**	**Missing**
FAMPMT	Yes	1	1	2
	No	1	N/A	4
	Missing	3	5	6

N/A = not applicable.

Both the likeness and logical constraints are written in such a way that they can be summarized using the five broad missingness patterns listed in Table E.100. For example, likeness constraint 4 states, "If recipient is missing months on welfare, then donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing)." This constraint does not apply to the first 255 missingness patterns, nor does it apply to the 64 missingness patterns in combination 6 where both FAMPMT and FAMSVC are missing. However, instead of carefully listing the exact missingness patterns to which each constraint applied, the constraints were written using conditions that allowed them to be applied only where necessary.

The few logical constraints are due to relationships between (1) FAMPMT, FAMSVC, and WELMOS; and (2) PINC1, FINC1, and FAMINC1. A key likeness constraint exploited the high degree of association among welfare payments, welfare services, food stamps, binary total income (at the personal and family level), and months on welfare. If a recipient required imputation for one or more of these six variables, but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables.

As stated in the introduction to Section 9.3, the variables for total personal income, total family income, and number of months on welfare have lower item response rates compared with most of the other variables that undergo imputation. The constraints are fairly complex and extensive, especially when several of the other dichotomous income variables are missing. Also,

[132] A reasonable alternative method that requires no assumption of independence would be to model FAMPMT and FAMSVC together as a categorical variable with four levels: both, only FAMPMT, only FAMSVC, and neither. Then, the probability of receiving welfare payments and/or other welfare services is simply the sum of the first three predicted means. The assumption of independence is certainly questionable.

when several variables are missing, it is difficult to find a donor on the first try because the donor must be close to the recipient for all predicted means corresponding to the missing variables.

9.3.1.6 Recodes for Additional Analyses

A dichotomous recoded income variable GOVTPROG indicated whether the respondent participated in any government assistance programs. It was created from four imputation-revised variables: family Supplemental Security Income (IRFAMSSI), family food stamps (IRFSTAMP), family welfare payments (IRFAMPMT), and family welfare services (IRFAMSVC). Although a variety of recoded variables were created, only GOVTPROG is described here because it was used as a covariate in subsequent health insurance imputation models. See Chapter 10 for details on the imputation of missing values in the health insurance variables.

9.3.2 Imputation for Finer Category Income Variables (Imputation Set 2)

Three income variables resulted from editing the questions in the finer category income phase: personal total income (PINC2), total family income if there are other family members (FINC2), and total family income (FAMINC2). These edited variables are described in Section 9.2.2. All three imputation-revised variables derived from their edited counterparts were created using the single RP/single PRD PMN type described in Chapter 3. The single PRD model was fit for total family income (FAMINC2), but the item nonrespondent also received values from the donor for PINC2 and FINC2 if those were missing. There were no deviations from this general approach.

9.3.2.1 Response Propensity Step

The response propensity models for the finer category income variables utilized the preliminary analysis weight, PANALWT. All respondents were in the domain for this imputation set. Item respondents were those with nonmissing values for PINC2 and FINC2. See Table D.58 in Appendix D for details of the covariates used in the RP models for these variables.

9.3.2.2 Prediction Step

Each of the three finer category income variables was ordinal with 29 levels. Each response category covers an interval of income, with levels as follows:

1 LESS THAN $1,000 (INCLUDING LOSS)
2 $1,000 - $1,999
3 $2,000 - $2,999
4 $3,000 - $3,999
5 $4,000 - $4,999
6 $5,000 - $5,999
7 $6,000 - $6,999
8 $7,000 - $7,999
9 $8,000 - $8,999
10 $9,000 - $9,999
11 $10,000 - $10,999

12	$11,000 - $11,999
13	$12,000 - $12,999
14	$13,000 - $13,999
15	$14,000 - $14,999
16	$15,000 - $15,999
17	$16,000 - $16,999
18	$17,000 - $17,999
19	$18,000 - $18,999
20	$19,000 - $19,999
21	$20,000 - $24,999
22	$25,000 - $29,999
23	$30,000 - $34,999
24	$35,000 - $39,999
25	$40,000 - $44,999
26	$45,000 - $49,999
27	$50,000 - $74,999
28	$75,000 - $99,999
29	$100,000 OR MORE

The FAMINC2 variable was modeled using the LIFEREG procedure in SAS/STAT® software.[133] This procedure was used for regression modeling of continuous nonnegative random variables, such as survival times and income, by fitting models that are sometimes referred to as "failure time models." This particular type of model, which was assumed for the response variable representing income, can be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is a vector of observed responses, $\mathbf{X}$ is the matrix of covariates, $\boldsymbol{\beta}$ is the parameter vector, and $\boldsymbol{\varepsilon}$ is a vector of error terms. In particular, the error terms are assumed to come from a known multivariate distribution, such as the logarithm of a three-parameter generalized gamma model, or a more common two-parameter distribution, such as gamma, Weibull, lognormal, or log-logistic. Although the underlying random variable $\mathbf{y}$ is assumed to be continuous, the LIFEREG procedure allows the variable to be reported in interval categories, consistent with the 29 NSDUH income intervals defined previously for these finer category income variables. The contribution of an individual with covariates in the matrix $\mathbf{X}$ to the overall likelihood is simply the probability mass assigned by the model to the interval (l, u) containing the actual continuous income for that individual. For this interval, l represents the lower bound and u represents the upper bound. This contribution has the form $F(u|\mathbf{X},\boldsymbol{\beta},\sigma^2) - F(l|\mathbf{X},\boldsymbol{\beta},\sigma^2)$, where F is a cumulative distribution function and σ^2 represents the variance of the individual responses. The LIFEREG procedure uses standard likelihood methods of inference and incorporates the survey weights.

LIFEREG allows several choices for the functional form of the parametric model that correspond to the error distribution, including the two-parameter log-logistic, lognormal, gamma,

[133] Details about the LIFEREG procedure are discussed in the *SAS/STAT 13.1 User's Guide* (SAS Institute Inc., 2013).

and Weibull and the three-parameter generalized gamma. Compared with the other models, the gamma distribution provided a better overall fit, as indicated by likelihood techniques. Because the three-parameter generalized gamma did not significantly improve on its two-parameter special cases, when using the likelihood ratio tests as criteria for comparison, it was decided to use a two-parameter model.

The predicted mean used in the subsequent hot-deck step was the term $\mathbf{X\beta}$, which was the predictive mean value. This value was a monotonic function of the conditional mean of the modeled income distribution at a given individual set of values of the regression covariates. Specifically, $\mathbf{X\beta}$ was a translation of the estimated mean of log income.

9.3.2.3 Hot-Deck Step

The hot-deck step for the finer category income variables is an example of univariate matching, multivariate assignment (see Table 3.4 in Chapter 3). The only predicted mean used in the hot-deck step was related to FAMINC2, but the recipient received values for PINC2 and FINC2 if either or both were missing. The imputation-revised version of FAMINC2 (IRFAMIN2) was created from the imputation-revised versions of PINC2 and FINC2, called IRPINC2 and IRFINC2, respectively. The constraints ensure consistency with existing information such as IRPINC1 and IRFINC1 from the preceding imputation set, and PINC2 and FINC2 if nonmissing.

The finer category income variables have among the lowest response rates of any NSDUH variables that undergo imputation. As a result of the absence of many constraints, the single predicted mean, and the large domain, most respondents are handled in the first attempt to find a donor.

9.3.2.4 Recodes for Additional Analyses

The recoded variable INCOME classified the families of respondents into four income levels: less than $20,000; $20,000 to $49,999; $50,000 to $74,999; and $75,000 or more. Another recoded variable (INCOME5) was created to take advantage of an extra level of income. This variable had five levels: the first three levels were equivalent to INCOME, but the last level of INCOME was separated into two levels: $75,000 to $99,999 and $100,000 or more. Both INCOME and INCOME5 were recodes of the variable IRFAMIN2. A variety of recoded variables were created but are not discussed in this report. However, as with GOVTPROG, the variable INCOME is discussed here because it was used as a covariate in subsequent health insurance models (see Chapter 10 for details on the imputation of missing values in the health insurance variables). INCOME5, which is currently used for special requests, also is discussed because it is similar to the INCOME variable.

10. Editing and Imputation for the NSDUH Health Insurance Variables

10.1 Introduction

This chapter discusses procedures for logically editing and imputing data from the interviewer-administered sections pertaining to health insurance coverage. Respondents (or other household members serving as proxies because they were better able to provide information about the respondent's health insurance coverage) were asked whether they were currently covered by different types of health insurance.[134] If private health insurance coverage was reported, respondents were asked whether that included coverage for substance abuse treatment or mental health services. Data also were collected on periods when respondents never had health insurance coverage, most recent coverage (if currently without health insurance), and reasons for losing health insurance coverage or for never having had coverage (if applicable).

For the 2012 National Survey on Drug Use and Health (NSDUH), as has been the case since 2002, missing values for the health insurance variables were replaced with valid values using two different methods: the "old" method and the "constituent variables" method. Both are predictive mean neighborhood (PMN) imputation methods. The old method imputes for three health insurance variables in a way that is consistent with iterations of the NSDUH questionnaire prior to 2002. The constituent variables method imputes for specific health insurance variables (e.g., Medicaid or Medicare coverage) and was first implemented for the health insurance module in 2002. As with other variable groups (i.e., demographics, drugs, etc.), the health insurance variables were imputed in sets, as follows:

- Imputation Set 1: IRINSUR, IRINSUR3, and IRPINSUR (Old Method: any health insurance and private health insurance)
- Imputation Set 2: IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT (Constituent Variables Method: types of health insurance)
- Imputation Set 3: IROTHHLT (Constituent Variables Method: any other health insurance)

Under the old method, the three base variables (INSUR, INSUR3, and PINSUR) that undergo imputation are simple recodes created from six "source" variables, each of which maps to a single question in the NSDUH. In the 2012 survey, the source variables had the same values as the raw variables created from the responses to the questions, except that missing values were replaced by standard NSDUH missing value codes (see Sections 2.2.1 and 2.4.3).

Under the constituent variables method, the six source variables are the same as those used in the old method, but they are not recoded and combined prior to being imputed. With the

[134] Reference is made only to "respondents" in the remainder of this chapter. However, readers are advised that the health insurance information for a respondent may have been provided by another adult household member who was serving as a proxy for the respondent because the proxy was considered to be better able to answer the health insurance questions for the respondent. See Chapter 8 for more information about proxy variables.

exception of MEDICAID and CHIPCOV, the variables remain separate to form the base variables for imputation sets 2 and 3.

Section 10.2.5 discusses the creation of base variables for imputation under the old method. These are the base variables used in imputation set 1. Section 10.2.6 discusses the creation of base variables for imputation under the constituent variables method. These are the base variables used in imputation sets 2 and 3.

10.2 Editing the Health Insurance Variables

Because the health insurance questions are in a noncore section of the interview (see Section 1.1), the content of these questions can change across years to improve the measurement of health insurance coverage. In practice, the content of this module has undergone relatively little change since the start of the new baseline in 2002. Documentation of historical changes to the health insurance questions prior to the new baseline in 2002 can be found in the report on editing of interviewer-administered data that was prepared for the 2011 Methodological Resource Book (Kroutil, Chien, Handley, & Bradshaw, 2013).

An important aspect of editing the health insurance variables consisted of assigning legitimate skip codes based on the skip logic in this section (Section 2.4.2). Figure 10.1 shows the logic for asking questions in the health insurance module. For example, if respondents answered "no" (where applicable) to questions QHI01 through QHI06 and then reported in QHI11 that they were not currently covered by any kind of health insurance (QHI11 = 2), legitimate skip codes were assigned to HLCNOTYR (anytime in the past 12 months that respondents were without health insurance, corresponding to question QHI13) and HLCNOTMO (number of months that respondents were without health insurance in the past 12 months, corresponding to question QHI14). Similarly, if respondents reported some type of current health insurance coverage in QHI01 through QHI06, edited variables corresponding to questions QHI15 through QHI18 were assigned legitimate skip codes (i.e., HLCLAST through HLNVSOR).

As was the case in prior years, question QHI18 (reasons that the respondent never had health insurance) was an "enter all that apply" question. Therefore, the edited variables corresponding to question QHI18 (HLNVCOST through HLNVSOR) were assigned a code of 1 (Response entered) if the corresponding response category was chosen from QHI18. The variables were assigned a code of 6 (Response not entered) if the corresponding response category was not chosen but at least one response had been entered in QHI18 (see Section 2.4.4).

Figure 10.1 Routing Logic for Respondents in the Health Insurance Module

*Respondents immediately go to the next module (i.e., without further health insurance questions being asked) if current health insurance coverage status is unknown.

10.2.1 Editing Because of Incorrect State Information

Question QHI02 asked whether respondents were currently covered by Medicaid, and question QHI02A asked whether respondents aged 12 to 19 were covered by the Children's Health Insurance Program (CHIP). The CAI program filled in names for State-specific Medicaid program or CHIP names to aid respondent identification. If the interviewer's report of the State where the dwelling unit was located (edited variable STATELOC) had been set to bad data because of inconsistencies in the State information for the respondent (see Section 4.2.3), then the edited variable CAIDCHIP for coverage by Medicaid or CHIP (see Section 10.2.2) was usually assigned a bad data code as well. The rationale for this edit was that the CAI logic would supply an incorrect name for the State's Medicaid program or CHIP if the State information that the interviewer reported was incorrect. Consequently, the respondent would be answering QHI02 or QHI02A based on a version of the question that did not correctly correspond to where the respondent would be eligible for publicly funded health insurance coverage. For example, if a respondent was aged 12 to 19 and was living in California (FIPE4 = 5), the respondent should have been asked in QHI02A whether he or she was covered by "Healthy Families." However, if a value of 6 had been entered in FIPE4 (i.e., for Colorado), the respondent would be asked whether he or she was covered by "Child Health Plan Plus, or CHP+."

An exception to this assignment of bad data codes concerned the special situation in which respondents were routed to questions QHI15 (time since the respondent last had health insurance) and QHI17 (main reason for losing health insurance coverage). If responses to QHI15 or QHI17 indicated that the respondent did not currently have (or never had) health insurance coverage, CAIDCHIP retained a code of 2 (i.e., "no"), even if STATELOC had been set to bad data, for consistency with information from QHI15 or QHI17 that the respondent was not currently covered by any type of health insurance.

10.2.2 Determination of Any Health Insurance Coverage

Determining whether respondents were currently covered by any type of health insurance was critical for determining the subsequent questions that respondents would be asked about periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, or reasons for never having health insurance. The following content issues have been relevant since the new NSDUH baseline in 2002 for determining whether respondents had health insurance coverage:

- As noted previously, respondents who were aged 12 to 19 are asked in question QHI02A whether they were covered by CHIP. Government experts in the health insurance field advised SAMHSA that it would be virtually impossible to produce separate estimates of coverage by the Medicaid program (question QHI02) and coverage by CHIP. For this reason, the variable CAIDCHIP has been created from responses to questions QHI02 and QHI02A.
- The State-specific Medicaid or CHIP names that are used to customize questions QHI02 and QHI02A may be updated for the CAI program across years to reflect any changes to these names at the State level.

- If STATELOC had a valid value (see Sections 4.2.3 and 10.2.1), CAIDCHIP was assigned a code of 1 (i.e., "yes") if an affirmative response occurred in either QHI02 or QHI02A (if applicable). CAIDCHIP was coded as 2 ("no") if QHI02 was answered as "no" and (a) QHI02A also was answered as "no" (for respondents who were aged 12 to 19) or (b) QHI02A had been legitimately skipped (for respondents aged 20 or older). Otherwise, CAIDCHIP was coded as 97 ("refused") if a code of 97 occurred in either QHI02 or QHI02A, or 94 ("don't know") if a code of 94 (and no code of 97) occurred in these items. Remaining cases that did not meet any of these criteria were coded as 98 (i.e., blank).
- Respondents who answered "no" to all questions about Medicare, Medicaid, CHIP (if applicable), military health coverage, and private health insurance were asked a follow-up question (QHI11) to determine if they were covered by *any* type of health insurance. Responses to this question were used to determine subsequent routing in the health insurance section depending on whether respondents currently had or did not have health insurance. The variable HLTINNOS was created from QHI11.
- A recoded "any health insurance" variable, ANYHLTIN, was created from responses to MEDICARE (from QHI01), CAIDCHIP (from QHI02 and QHI02A), CHAMPUS (from QHI03), PRVHLTIN (from QHI06), and HLTINNOS (from QHI11).
  - If any affirmative response was reported in any of the above variables, ANYHLTIN was coded as 1 ("yes").
  - Otherwise, if HLTINNOS (from QHI11) had been answered as "no" (and by definition, preceding questions had been answered as "no"), ANYHLTIN was coded as 2 ("no").
  - If ANYHLTIN was not already coded as 1 or 2, it was coded as 97 ("refused") or 94 ("don't know"), as follows: (a) if a code of 97 (i.e., "refused") occurred in any of the above health insurance items, ANYHLTIN was coded as 97; or (b) ANYHLTIN was coded as 94 (i.e., "don't know") if a code of 94 (but no code of 97) occurred in the above items.
  - For remaining cases (e.g., if variables had been set to bad data, or a breakoff had occurred), ANYHLTIN retained a code of 98 (OTHER MISSING).

In addition, because the names for Medicare and Medicaid sound similar, follow-up questions have been included in the health insurance section of the interview since 2003 for respondents who were (a) under the age of 65 and reported being covered by Medicare (which is for persons aged 65 or older or certain disabled persons), or (b) aged 65 or older and reported being covered by Medicaid (which is for low-income persons or disabled persons). In these situations, respondents were asked to confirm their previous answer. If respondents indicated on follow-up that their previous answer to Medicare or Medicaid coverage was incorrect, respondents were determined not to be covered by that particular type of health insurance. However, if respondents did not know or refused to confirm whether their previous answer to the Medicare or Medicaid coverage question was correct, the editing procedures continued to classify them as having that type of coverage. Stated another way, the only response on follow-up that overruled a previous report of coverage by Medicare or Medicaid was for respondents to indicate that they did not have that type of coverage.

Despite these changes in 2003 for questions about Medicare or Medicaid coverage, the names for the edited variables did not change. The edited variable for Medicare coverage retained the name MEDICARE. The edited variable for coverage by Medicaid or CHIP retained the name CAIDCHIP (see previous discussion in this section).

10.2.3 Editing of Health Insurance Follow-Up Variables Based on Current Coverage

Information from ANYHLTIN and a preliminary edited variable SKHLCCOV[135] were used to determine whether variables pertaining to periods without health insurance coverage in the past 12 months, most recent health insurance coverage (if not currently covered), reasons for losing health insurance, and reasons for never having health insurance should be assigned "legitimate skip" codes (Section 2.4.2).

- If respondents reported some health insurance coverage in QHI01 to QHI11, legitimate skip codes were assigned to the edited variables corresponding to questions QHI15 to QHI18. Questions QHI15 to QHI18 applied to respondents who currently were without coverage and pertained to situations in which respondents formerly or never had coverage.
- If respondents currently were without any health insurance coverage (i.e., applicable questions in QHI01 to QHI11 all were answered as "no," legitimate skip codes were assigned to the edited variables corresponding to questions QHI13 and QHI14. Questions QHI13 and QHI14 applied to respondents who currently had coverage and pertained to periods when these respondents were without coverage in the past 12 months.
- If it was unknown from questions QHI01 to QHI11 whether or not respondents currently had health insurance coverage, the CAI program skipped respondents out of all remaining health insurance questions. The edited variables corresponding to questions QHI13 to QHI18 had missing values.

10.2.4 Miscellaneous Issues for Editing the Health Insurance Variables

Table 10.1 discusses edit specifications for the health insurance variables. For example, the data could indicate that respondents were covered currently by Medicare, Medicaid, CHIP (for respondents who were aged 12 to 19), some type of military health coverage (e.g., CHAMPUS or the VA), or private health insurance. If respondents were reported to have been *currently* covered by all of the types of insurance they were asked about, a flag was set and included on the data file. The original data were retained, but this flag was designed to alert analysts to the presence of this unlikely data pattern.

[135] SKHLCCOV was a precursor to the final variable ANYHLTIN but was used in some places in the editing program to determine whether respondents were legitimately skipped out of certain follow-up questions about health insurance coverage. Both variables indicate whether respondents currently had health insurance coverage. SKHLCCOV was coded as 3 if respondents reported in question QHI02A that they were covered by CHIP but they did not report coverage by Medicaid, Medicare, military health insurance (e.g., CHAMPUS), or private health insurance. These values of 3 in SKHLCCOV were recoded to 1 in ANYHLTIN.

In addition, the only types of current health insurance coverage that were asked about in 1999 were Medicare, Medicaid, some type of military health coverage, or private health insurance. Therefore, a second flag was set for comparability with a similar flag set in the 1999 data. This second flag indicated when respondents reported that they were covered currently by all four of these types of health insurance that were asked about in 1999, even if they did not report being covered by CHIP (if aged 12 to 19) or they were aged 20 or older and were skipped out of question QHI02A.

Table 10.1 Edits Pertaining to the Health Insurance Section

Issue	Edits Implemented
The respondent (R) reported being currently covered by Medicare, Medicaid, Children's Health Insurance Program (CHIP) (if aged 12 to 19), military coverage, and private health insurance.	A flag (HLCALLFG) was provided to indicate that this pattern occurred, but no further editing was done to the data.
The R reported being currently covered by Medicare, Medicaid, military coverage, and private health insurance, the only types of current coverage that were asked about in 1999.	A flag (HLCALL99) was provided to indicate that this pattern occurred, but no further editing was done to the responses. This HLCALL99 variable was comparable to the HLCALLFG variable in 1999.
The R's only indication of current health insurance coverage came from reports of coverage by Medicaid or CHIP, but the State location variable STATELOC had been set to bad data.	Nonblank values in the variables pertaining to any period in the past 12 months when the R was without health insurance (HLCNOTYR, corresponding to question QHI13) and the number of months that the R was without health insurance in the past 12 months (HLCNOTMO, corresponding to question QHI14) were replaced with bad data codes. This edit was not done if the R indicated current coverage by Medicare, the military, or private health insurance.
The R had some indication of current coverage from at least one of the five sources of insurance listed above. However, the R also was reported to have had a period in the past 12 months when he or she was without health insurance. Further, it was reported that the R had been without health insurance for 12 of those months.	No editing was done when this pattern occurred. The rationale for not doing any editing was that the R may just recently have gotten insurance or have become qualified for insurance.
The R had no indication of current coverage from any of the five sources of insurance listed above. If the R (or proxy) answered "don't know" or "refused" when asked when the R last had coverage, the R was routed to questions about what coverage the R last had, and why the R lost health insurance coverage. That is, the skip logic assumed that the R had some prior history of coverage, but that may not necessarily have been the case.	If the R was reported to have had previously some form of health insurance or medical coverage, or if some reason was given why the R lost insurance coverage, legitimate skip codes were assigned to the variables pertaining to reasons that the R *never* had coverage. That is, the implicit assumption made in the computer-assisted interviewing (CAI) skip logic was verified by an answer indicating some prior history of health insurance coverage. However, if nothing was reported to indicate that the R previously had health insurance, the skipped variables pertaining to reasons for never having had insurance retained codes of blank.
The R was male but reported in QHI17 that he lost health insurance coverage because he "received Medicaid or medical insurance only while pregnant."	The edited variable HLLOSRSN (corresponding to QHI17) was set to bad data.

10.2.5 Editing Any Health Insurance and Private Health Insurance Variables, Imputation Set 1 (Old Method)

Under the old method, the three base variables (INSUR, INSUR3, and PINSUR) were created from the six source variables shown in Table 10.2. INSUR and INSUR3 indicated whether the respondent had any health insurance, and PINSUR indicated whether the respondent had any private health insurance. INSUR, which was created to maintain consistency with the 1999 survey, was coded as "yes" if any of MEDICARE, MEDICAID, CHAMPUS, or PRVHLTIN were coded as "yes"; it was coded as "no" if all four were coded as "no." In 2001, the questions associated with CHIPCOV and HLTINNOS were added to the questionnaire. The variable INSUR3 was created beginning with the 2001 survey to incorporate the addition of these two new questionnaire items. INSUR3 was coded as "yes" if any one of the six variables listed in Table 10.2 were coded as "yes"; it was coded as "no" if all six variables were coded as "no."[136] PINSUR was a direct recode of PRVHLTIN.

Table 10.2 Mapping of Raw Health Insurance Variables to Base Variables for Imputation Set 1 (Old Method)

Question Number[1]	Question Text[2]	Source Variable[3]	Used To Create INSUR	Used To Create INSUR3	Used To Create PINSUR
QHI01, QHI01v	Is the respondent covered by Medicare?	MEDICARE (1 = yes, 2 = no)	Yes	Yes	No
QHI02, QHI02v	Is the respondent covered by Medicaid or Medical Assistance?	MEDICAID (1 = yes, 2 = no)	Yes	Yes	No
QHI02A	Is the respondent currently covered by a Children's Health Insurance Program operated by your State of residence?[4]	CHIPCOV (1 = yes, 2 = no)	No	Yes	No
QHI03	Is the respondent currently covered by CHAMPUS or TRICARE, CHAMPVA, the VA, or military health care?	CHAMPUS (1 = yes, 2 = no)	Yes	Yes	No
QHI06	Is the respondent currently covered by private health insurance?	PRVHLTIN (1 = yes, 2 = no)	Yes	Yes	Yes
QHI11	Is the respondent currently covered by any kind of health insurance, that is, any policy or program that provides or pays for medical care?	HLTINNOS (1 = yes, 2 = no, 99 = legitimate skip[5])	No	Yes	No

[1] The "v" questions were asked to verify the answer given in the previous question for respondents who were younger than 65 and a Medicare recipient or older than 65 and a Medicaid recipient.
[2] The questions provided in this table are abbreviated versions of those given in the questionnaire.
[3] Missing values in these edited values were represented by standard missing value codes. CHIPCOV was replaced in the final analytic file by CAIDCHIP, a combination of MEDICAID and CHIPCOV. See Section 10.2.2 for details.
[4] The questionnaire did not ask the question exactly in this way. It identified the specific program, depending upon the State of residence entered by the respondent and was asked only of respondents aged 12 to 19.
[5] Respondents were assigned a legitimate skip for HLTINNOS if they answered "yes" or gave no answer to at least one of the other health insurance questions.

[136] In the 2000 survey, the variable INSUR2 was created to take advantage of the additional information provided by questions that did not exist in the 1999 questionnaire. However, because these additional questions were either replaced or reworded in later surveys, the variable INSUR2 has not been used in the surveys since 2000.

10.2.6 Editing Types of Health Insurance and Any Other Health Insurance Variables, Imputation Sets 2 and 3 (Constituent Variables Method)

Under the constituent variables method, the editing process for the health insurance imputation process combined the source variables (MEDICAID and CHIPCOV) to create the variable CAIDCHIP (whether someone was covered by Medicaid or one of the State children's health insurance programs). This CAIDCHIP variable and all the other source variables in Table 10.2, except HLTINNOS, were used directly as base variables for imputation.

A respondent was routed to QHI11 (whether the respondent was covered by any kind of health insurance at the time of the survey) if they answered "no" to all the other health insurance questions. All other respondents were given a legitimate skip value to the variable HLTINNOS, as shown in Table 10.2. Therefore, it was possible that the imputation-revised versions of the four specific health insurance variables would all have had a value of "no," and the value of HLTINNOS would have been a legitimate skip, if one or more of the "no" values was imputed. In this instance, another variable was needed to reflect the fact that a respondent could have had a valid "yes" or "no" imputed value for any other health insurance, even though the respondent was never asked QHI11 and was assigned a legitimate skip code. The ANYOTHER variable was created using the preliminary edited variable, SKHLCCOV (Section 10.2.3), which indicated whether a respondent was covered by any health insurance. SKHLCCOV and ANYOTHER were defined as follows:

SKHLCCOV =

- 1 (or 3[137]) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, PRVHLTIN = 1, or HLTINNOS = 1; else
- 2 if CAIDCHIP = 2, MEDICARE = 2, CHAMPUS = 2, PRVHLTIN = 2, and HLTINNOS = 2; else
- missing value code if the nonmissing values of CAIDCHIP, MEDICARE, CHAMPUS, PRVHLTIN, and HLTINNOS are all "2," and at least one of these variables had a missing response.

ANYOTHER =

- legitimate skip code (99) if CAIDCHIP = 1, MEDICARE = 1, CHAMPUS = 1, or PRVHLTIN = 1; else
- SKHLCCOV.

10.3 Imputation for the Health Insurance Variables

For the 2012 NSDUH, missing values for the health insurance variables were replaced with valid values using two different methods: the "old" method and the "constituent variables" method. Under the old method, three health insurance variables comprise imputation set 1. The

[137] See Section 10.2.3 for discussion of the assignment of a code of 3 to SKHLCCOV. Respondents with SKHLCCOV = 3 were treated in the same manner as those with SKHLCCOV = 1.

first and second variables, IRINSUR and IRINSUR3, were both indicators of "any health insurance" coverage. The third variable, IRPINSUR, was simply an imputation-revised version of the edited "private health insurance" variable, PRVHLTIN, detailed in Section 10.2.2. These different versions of health insurance coverage indicators were created because the question set changed between 1999 and 2001. See Section 10.2.5 for more detail on the creation of these variables. Section 10.3.1 describes the creation of the three imputation-revised health insurance variables for imputation set 1.

Under the constituent variables method,[138] indicator variables for more specific types of health insurance (i.e., coverage by Medicare or by Medicaid) were imputed in two additional sets. Imputation set 2 included the variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. These are indicators of coverage for Medicaid/CHIP, Medicare, CHAMPUS or similar coverage for military personnel, and private health insurance, respectively. Section 10.3.2 describes the creation of the four imputation-revised health insurance variables that comprise imputation set 2. Within imputation set 3, the "any other health insurance" variable, IROTHHLT, was created (see Section 10.3.3). Together, these five constituent imputation-revised health insurance variables were then used to create a third indicator of "any health insurance" coverage, IRINSUR4.

Regardless of whether the final health insurance variables were derived by the old method or the constituent variables method, imputations were performed using the same methodology, the PMN method. The health insurance variables that undergo imputation tend to have item response rates of more than 99 percent. See Chapter 3 for details on PMN; see Table A.27 in Appendix A for details on rates of missingness for the health insurance variables.

10.3.1 Any Health Insurance and Private Health Insurance Variables, Imputation Set 1 (Old Method)

The imputation type used for the three health insurance variables imputed under the old method differs from the three types described in Chapter 3, but it is most similar to single response propensity (RP)/multiple prediction (PRD). Even though there were three variables in the set, only one RP model was fit and only two PRD models were fit, but there was no provisional hot-deck step. Missing data were filled in for all three variables in a single, final hot-deck step. Since the sole purpose of the old method was to maintain consistency with earlier NSDUHs, the method has not been updated or revised to make it more consistent with the approaches used for other variables. For example, SAS procedures were used to fit the PRD models instead of SUDAAN procedures to maintain consistency with pre-2002 data processing. These are among the few PRD models in the NSDUH imputation process that continue to use SAS instead of SUDAAN. For a complete list of PRD models and the types of procedures used to fit them, see Table 3.2.

[138] The constituent variables method varies slightly from other methods used to impute variables for the NSDUH. First, it uses some uncommon constraints and covariates. For example, from the core demographics module, the SERVICE variable, an indicator of service in the United States armed forces, was used both as a covariate and in a likeness constraint related to CHAMPUS. Second, age groups were created to ensure reasonable domain sizes for both modeling and hot-deck steps. Finally, provisional imputation steps in imputation set 2 were not actually hot-deck steps but simple stochastic imputations based on predicted means, and in the final hot-deck step for these variables, different constraints were applied to different age groups.

The sequence of imputation is not important for health insurance imputation under the old method, because no provisional hot-deck step is involved. In fact, the two PRD models can be run in parallel, though this is not done in practice. Rather, the PRD model associated with INSUR3 was fit before the PRD model associated with PINSUR was fit, and no PRD model was fit for INSUR. Since the response rates were high for these two variables, the single RP/multiple PRD type of PMN was appropriate for this imputation set.

10.3.1.1 Response Propensity Step

The response propensity models for imputation set 1 utilized the final analysis weight, ANALWT, as an input. All respondents were in the domain for the health insurance variables imputed under the old method. No respondents received skip codes for the final imputation-revised variables IRINSUR, IRINSUR3, and IRPINSUR. For the single RP model, a domain member was considered an item respondent if INSUR3 and PINSUR were nonmissing. See Tables D.65 and D.66 in Appendix D for details of the covariates used in the RP models for these variables.

10.3.1.2 First Prediction Step (INSUR3)

INSUR3 was a dichotomous variable; after imputation, its only values are "yes" and "no." Therefore, it was modeled using logistic regression and the adjusted weights that are outputs of the RP step. This PRD model was unusual in that the LOGISTIC procedure in SAS was used instead of the RLOGIST procedure in SAS-callable SUDAAN. The old method was implemented only to maintain consistency with earlier NSDUHs, in which the SAS procedure was used instead of the corresponding SUDAAN procedure.[139] The single predicted mean used in the later hot-deck step was the predicted probability that the respondent received overall health insurance, as defined by INSUR3.

10.3.1.3 Second Prediction Step (PINSUR)

The response indicator used in the single RP model was used in this PRD model, as were the adjusted weights. However, the domain indicator was different. Only respondents with INSUR3 = Yes were in the domain.

Like INSUR3, PINSUR was a yes/no variable, and it was modeled using the LOGISTIC procedure in SAS. However, since the domain was restricted as described in the preceding section, the single predicted mean used in the later hot-deck step was conditional. In particular, it was the predicted probability that the respondent received private health insurance as defined by PINSUR, given that the respondent was known to have received overall health insurance as defined by INSUR3.

10.3.1.4 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the final hot-deck step for old method health insurance are available in Appendix E. This section explains the more general ideas behind the hot-deck step applied to these variables.

[139] See Section 3.3.2 for a brief comparison of SAS and SUDAAN with regard to prediction models.

There are only six possible missingness patterns for old method health insurance. The three base variables are INSUR, INSUR3, and PINSUR, each of which can have one of three values: yes, no, or missing. This leads to $3 \times 3 \times 3 = 27$ possible combinations. However, some of these 27 combinations do not require imputation (because none of the three are missing), and there are logical relationships that render other combinations impossible. A glance at how the variables are created reveals that PINSUR is "nested" within INSUR, and INSUR is nested within INSUR3. Because of this nesting, the following logical relationships hold:

- PINSUR = yes implies that INSUR = yes and INSUR3 = yes.
- INSUR = yes implies that INSUR3 = yes.
- INSUR3 = no implies that INSUR = no and PINSUR = no.
- INSUR = no implies that PINSUR = no.
- INSUR3 = missing implies that INSUR ≠ yes and PINSUR ≠ yes.
- INSUR = missing implies that PINSUR ≠ yes.

The only remaining possibilities are the six missingness patterns.

The logical constraints preserve the logical relationships between the three variables. The likeness constraints are based only on the predictive means and the AGE variable. The predictive mean vectors are, where appropriate, conditioned on what is known; since the predicted mean associated with PINSUR is conditional, the conditional part has to be undone for missingness patterns where INSUR3 is missing. Finally, since (1) the item response rate is high for these variables, (2) the domain includes all unit respondents, and (3) the constraints are not very restrictive, the vast majority of item nonrespondent cases found donors on the first attempt.

10.3.2 Types of Health Insurance Variables, Imputation Set 2 (Constituent Variables Method)

The first four imputation-revised constituent variables were created using the single RP/multiple PRD PMN type. There were no noteworthy deviations from this general approach. The four imputation-revised variables thus created were yes/no indicators of whether the respondent had health insurance from Medicaid or a State children's health insurance program (IRMCDCHP), Medicare (IRMEDICR), CHAMPUS (IRCHMPUS), or private health insurance (IRPRVHLT). The health insurance indicators were imputed in the following order: CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. Since the response rates were high for these variables, the single RP/multiple PRD type of PMN was appropriate for this imputation set.

10.3.2.1 Response Propensity Step

The response propensity models for imputation set 2 utilized the final analysis weight, ANALWT. All respondents were in the domain for this stage of the constituent variables method. No respondents received skip codes for the final imputation-revised variables IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. For the RP model, a domain member was considered an item respondent if all four base variables (CAIDCHIP, MEDICARE,

CHAMPUS, and PRVHLTIN) were nonmissing. See Table D.60 in Appendix D for details of the covariates used in the RP models for these variables.

10.3.2.2 First Prediction Step (CAIDCHIP)

CAIDCHIP was a dichotomous variable; after imputation, its only values were "yes" and "no." Therefore, it was modeled using logistic regression and, in particular, using the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean used in the later hot-deck step was the predicted probability that the respondent received Medicaid or was covered by a State children's health insurance program.

10.3.2.3 First Provisional Stochastic Imputation Step (CAIDCHIP)

The provisional imputation step used for CAIDCHIP was not a hot-deck step, but a stochastic imputation. Each recipient was assigned a "yes" response for the intermediate imputation-revised variable INTCDCH with probability equal to the predicted mean from the model. This method was based on the idea of a centered PMN, as described in Singh, Grau, and Folsom (2004). As applied to single dichotomous variables like CAIDCHIP, the centered PMN approach reduces to a simple stochastic imputation that requires no donors and no constraints. This approach to imputation for categorical variables is further explored in the PMN imputation evaluation report (Ault et al., in press).

10.3.2.4 Analogous Prediction and Provisional Hot-Deck Imputation Steps (MEDICARE, CHAMPUS, and PRVHLTIN)

PRD models were fit for MEDICARE, CHAMPUS, and PRVHLTIN in the same manner as for CAIDCHIP, as described above, with a few notable deviations:

- For the MEDICARE variable, a single PRD model was fit for the 18-to-64 age group. This was done because (1) only a small proportion of respondents in these age groups had Medicare, particularly for the 18-to-25 age group; and (2) a respondent of working age could have received Medicare only if he or she was not working because of disability. This was true regardless of whether the respondent was aged 18 to 25 or 26 to 64.
- The CHAMPUS PRD models used a covariate that underwent a simplified imputation procedure. For respondents aged 18 or older, models included an indicator of whether the respondent had ever been in the military service, designated by an imputation-revised version of the edited variable SERVICE. The variable SERVICE generally had a very low level of missingness (one missing value in the 2012 survey). Because covariates in these models were not supposed to have any missing values, the missing values in the SERVICE variable were randomly imputed as "yes" responses if the random number was greater than the mean value of SERVICE across all the other respondents, and imputed as "no" otherwise.
- PRVHLTIN did not undergo a provisional imputation step since it was the last variable in the set.

10.3.2.5 Final Hot-Deck Step

Details on the missingness patterns, constraints, and predictive mean vectors for the final hot-deck step of this first stage of imputation under the constituent variables method are available in Appendix E. This section explains the general ideas behind the hot-deck step applied to these variables.

There are fifteen missingness patterns for the first stage of constituent variables method health insurance imputation. Each of the four variables in the set can be missing or nonmissing. This leads to $2 \times 2 \times 2 \times 2 = 16$ possible combinations, but the combination where none are missing is one that does not require imputation. No logical constraints were applied, since there were no logical relationships among the variables (or between the variables and any other NSDUH variables). The predictive mean vectors were also made simple because of the absence of logical relationships. Nonetheless, the likeness constraints were fairly sophisticated and sometimes varied across age groups, exploiting auxiliary variables from the demographics and income modules. The SERVICE variable was also used, but not an imputation-revised version of it; the constraint only applied when the item nonrespondent had a nonmissing value. As had been the case for variables imputed under the old method, the vast majority of item nonrespondents in imputation set 2 are typically handled on the first attempt to find a donor because (1) the item response rates are high, (2) the domain includes all unit respondents, and (3) the constraints are not very restrictive.

10.3.3 Any Other Health Insurance Variable, Imputation Set 3 (Constituent Variables Method)

The final imputation-revised constituent variable, IROTHHLT, indicated whether respondents had any type of health insurance, even though they reported or were imputed to have none of the four types of specific health insurance, as recorded by IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT. IROTHHLT was created from the base variable ANYOTHER. The PMN imputation type used for this variable was single RP/single PRD, and there were no noteworthy deviations from this general approach.

For this second stage under the constituent variables method, the age groups were 12 to 17, 18 to 25, and 26 or older. Three age groups were used instead of four because of the small number of respondents who would have otherwise comprised the domain for the 65-or-older age group.

10.3.3.1 Response Propensity Step

The response propensity models for imputation set 3 utilized the final analysis weight, ANALWT. The domain for the ANYOTHER variable included respondents who had either a reported or imputed "no" value to all four imputation-revised specific health insurance variables from the first stage (IRMCDCHP, IRMEDICR, IRCHMPUS, and IRPRVHLT). For a domain member to be considered an item respondent, he or she had to have complete data for the variable ANYOTHER. See Table D.60 in Appendix D for details of the covariates used in the RP model for this variable.

10.3.3.2 Prediction Step

Like all the other health insurance variables which undergo imputation, ANYOTHER was a dichotomous variable; after imputation, its only values are "Yes" and "No." Like the variables in the first stage, it was modeled using logistic regression, as implemented by the RLOGIST procedure in SAS-callable SUDAAN. The single predicted mean that was output by the RP step and used in the later hot-deck step was the predicted probability that the respondent received other health insurance, given that he or she was not covered by Medicaid/CHIP, Medicare, CHAMPUS, or private health insurance.

10.3.3.3 Hot-Deck Step

The hot-deck step for the ANYOTHER variable was the simplest one used in the NSDUH. There were no logical constraints, and the only likeness constraint was the delta constraint. The predictive mean vector was actually a scalar. Approximately 1 percent of the records in the domain underwent imputation, and nearly every item nonrespondent was handled on the first attempt to find a donor.

10.3.3.4 Recodes for Additional Analyses

The overall health insurance variable associated with the constituent variables method, IRINSUR4, was created by combining IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT. If a respondent had a reported or imputed "yes" value for any of these five variables, the respondent was considered to have health insurance. Otherwise, he or she did not have health insurance according to the constituent variables method. Though IRINSUR4 was technically a recoded variable created from other variables, an imputation indicator (IIINSUR4) was nevertheless created. IIINSUR4 was set to "1" if the respondent had a reported "yes" value for any of the five constituent health insurance variables or a reported "no" for all five of them; and "3" otherwise.

11. Editing and Imputation for the NSDUH Pair Variables

11.1 Introduction

In each household selected for the National Survey on Drug Use and Health (NSDUH), zero, one, or two household members are selected for interviewing. When two members of the same household are selected, the data for the responding pair (i.e., pair data) are used to study outcome variables based on the relationship between those household members (i.e., the pair relationship). For these analyses, the outcome variables can be at either the person level or the pair level. The most common type of analysis is the person-level analysis, where the inferential population is defined by one of the pair members (called the focus pair member). An example of an outcome at the person level is the proportion of children who use drugs and whose parents report talking to them about drugs, where the focus is on the child in a parent-child pair. By contrast, an example at the pair level is parent-child drug behavior for all possible parent-child pairs (within the child's age group). This chapter describes the techniques used to edit and impute the pair variables from the household roster (hereafter, "pair variables") for the 2012 NSDUH.[140] The variables described in this chapter can be divided into three stages:

- Stage One: Pair Relationships,
- Stage Two: Multiplicity Counts, and
- Stage Three: Household-Level Person Counts.

11.1.1 Pair Relationship Variable

The pair relationship variables are derived from the household composition (roster) variables, as described in Chapter 8 of this report. These variables include the edited and imputation-revised pair relationship variables PAIRREL[141] and IRPRREL, respectively, as well as the imputation indicator IIPRREL, which summarizes how the data in IRPRREL were obtained. In addition to these variables, the quality-of-match indicator RELMATCH and the pair indicator PAIRMEM (whether a respondent was part of a respondent pair) were created. Finally, four additional variables were created to aid in pair analyses: PRNTIND, AGEOTHER, SEXOTHER, and PAIRID. The variable PRNTIND identified whether the respondent was a parent in a parent-child relationship; AGEOTHER contained the age of the other respondent in the pair; SEXOTHER contained the gender of the other respondent in the pair; and PAIRID contained the questionnaire ID (QUESTID) of the other pair member.

[140] Edits to the variables for the parenting experiences module, which also involves selection of a pair, are discussed in Section 7.4.12 in Chapter 7. The parenting experiences module was administered if (1) an adult and an adolescent pair were selected for an interview, and (2) the adult was the parent of the adolescent who also was selected for an interview.

[141] The levels of PAIRREL are provided in Table 11.3.

11.1.2 Multiplicity Count Variables

"Multiplicity" is the term used to describe the complication that arises in the analysis of pair data in which the analysis is at the person level for a given pair domain, and several pairs in the household could be associated with the same person. For analyses at the pair level, the pair domain is completely and uniquely defined by the pair relationship. For example, to tabulate the number of sibling-sibling pairs where both siblings have used marijuana, it is necessary to know only whether a pair of respondents contains two siblings. By contrast, for person-level analyses, the pair domain depends upon which pair member is the focus. In this case, to analyze the influence that older siblings have on younger siblings in terms of drug use, it is necessary to know which pair member is the older sibling. The multiplicity count is a count of the number of pairs in the household that can be associated with the person of focus because—to continue the example above—a child may have more than one older sibling. If the multiplicity is not considered, then persons of focus who are members of more pairs receive too large a weight (i.e., they are counted too many times in the target population), and persons of focus who are members of fewer pairs receive too small a weight. Adjustment for multiplicity was not done a priori because that would lead to the addition of 12 weights to the file, one for each of the 12 possible pair relationships that require this adjustment (e.g., excluding the spouse-spouse and spouse-spouse with children relationships). Note that the multiplicity problem does not arise if there is only one inclusion possibility (e.g., a single-parent household, if the child is the focus) or if the analysis is a pair-level analysis (e.g., parent-child pair drug behavior).

To illustrate how multiplicities appear in the definitions of parameters and estimates, consider estimating the total number of children who used drugs in the past year, where a parent reported talking to them about drugs. Let $y_{hip}(d)$ be defined as the drug-related behavior outcome for pair *p* containing the individual *i* belonging to domain *d* in household *h*. Now, for the population of all individuals who belong to the domain *d*, the total parameter is defined as

$$\tau_y(d) = \sum_{h=1}^{H} \sum_{i=1}^{N_h(d)} \sum_{p=1}^{M_{hi}(d)} \frac{y_{hip}(d)}{M_{hi}(d)}$$

(i.e., total of averages over pairs (*p*) associated with the individual *i*, over all *i* in domain *d* and in the household *h*) (Chromy & Singh, 2001). Here $M_{hi}(d)$ denotes the multiplicity (i.e., the number of pairs associated with the individual *i* in domain *d*), and $N_h(d)$ can be thought of as the multiplicity count for the household *h* (i.e., the number of persons in the household that are in domain *d*). For the sake of simplicity, the weights are not shown in the above estimator.

Multiplicity count variables were only created for specific relationships of interest. These variables are listed in Table 11.1. With the exception of the spouse-spouse multiplicity variables listed at the bottom of this table, these variables all underwent imputation.

Table 11.1 Edited and Imputation-Revised Multiplicity Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of parents of a child aged 12-14 who is a member of a parent-child pair	MCPCC14	IRMPCC14
Number of children aged 12-14 belonging to a parent who is a member of a parent-child pair	MCPCP14	IRMPCP14
Number of parents of a child aged 15-17 who is a member of a parent-child pair	MCPCC57	IRMPCC57
Number of children aged 15-17 belonging to a parent who is a member of a parent-child pair	MCPCP57	IRMPCP57
Number of parents of a child aged 12-17 who is a member of a parent-child pair	MCPCC17	IRMPCC17
Number of children aged 12-17 belonging to a parent who is a member of a parent-child pair	MCPCP17	IRMPCP17
Number of parents of a child aged 12-20 who is a member of a parent-child pair	MCPCC20	IRMPCC20
Number of children aged 12-20 belonging to a parent who is a member of a parent-child pair	MCPCP20	IRMPCP20
Number of siblings aged 12-14 for a respondent aged 15-17 who is a member of a sibling-sibling pair	MCS1417	IRMS1417
Number of siblings aged 15-17 for a respondent aged 12-14 who is a member of a sibling-sibling pair	MCS1714	IRMS1714
Number of siblings aged 12-17 for a respondent aged 18-25 who is a member of a sibling-sibling pair	MCS1725	IRMS1725
Number of siblings aged 18-25 for a respondent aged 12-17 who is a member of a sibling-sibling pair	MCS2517	IRMS2517
Pair relationship is spouse-spouse with no children younger than 18	MCSPSP	N/A
Pair relationship is spouse-spouse with children younger than 18	MCSPSPWC	N/A

N/A = not applicable.

11.1.3 Household-Level Person Count Variables

Whereas the multiplicity count variable is the number of pairs in the household that can be associated with the focus pair member, the household-level person count variable is the number of persons of focus in the household for a given pair domain, provided such a pair domain existed in the household, regardless of which pair (or whether a pair) was actually selected. For example, if two parents were in the household with three children aged 12 to 14, the household person count for the *parent* focus of the parent-child (12-14) domain would be 2, and the household person count for the *child* focus would be 3. If the parents indicated that they had a spousal relationship, the household person count for the spouse-spouse with children domain would be 2. The rest of the household counts would be zero (even the sibling focus) because they do not fall in the age groups of the domains that are edited and imputed; that is, multiplicity counts were not created for siblings that were both aged 12 to 14. Also, note that household

person counts for all domains are calculated for every respondent, even when only one respondent was selected in the household or when a selected pair did not fall in a particular domain. The household-level person count variables are listed in Table 11.2, and all underwent imputation.

Table 11.2 Edited and Imputation-Revised Household-Level Person Count Variables

Variable Description	Edited Variable Name	Imputation-Revised Variable Name
Number of children aged 12-14 in the household with at least one parent living with them	HCPCC14	IRHPCC14
Number of parents in the household with at least one child aged 12-14 living with them	HCPCP14	IRHPCP14
Number of children aged 12-17 in the household with at least one parent living with them	HCPCC17	IRHPCC17
Number of parents in the household with at least one child aged 12-17 living with them	HCPCP17	IRHPCP17
Number of children aged 12-20 in the household with at least one parent living with them	HCPCC20	IRHPCC20
Number of parents in the household with at least one child aged 12-20 living with them	HCPCP20	IRHPCP20
Number of household members aged 15-17 with a sibling aged 12-14 living with them	HCS1417	IRHS1417
Number of household members aged 18-25 with a sibling aged 12-17 living with them	HCS1725	IRHS1725
Number of spouse-spouse pairs without children younger than 18	HCSPSP	IRHCSPSP
Number of spouse-spouse pairs with children younger than 18	HCSPSPWC	IRHCSPWC

11.1.4 Staged Variable Processing

The creation of the edited and imputation-revised pair variables was conducted in three stages because the variables from earlier stages were needed for the creation of variables in later stages. Stage one consisted of the creation and imputation of the variables that identify the pair relationships. The multiplicity and household-level person counts were created and imputed in stages two and three, respectively. Missing values in all three stages were imputed using the predictive mean neighborhood (PMN) imputation procedure, which uses predicted means from models to find donors in a nearest neighbor hot deck. Chapter 3 of this report provides background information about imputation in general (including hot-deck imputation) and details about the PMN methodology in particular.

Though this chapter presents first the editing procedures applied to the variables in each stage and then the imputation procedures for each stage, it is important to note that the actual order of processing was by stage; that is, both editing and imputation were completed for the variables in each stage before moving on to the next stage.

11.2 Stage One Editing: Pair Relationships

11.2.1 Editing the Household Roster of Each Pair Member

Prior to identifying the relationships between selected pair members, a key step is editing the questionnaire household rosters for each pair member. This involves identifying situations where the relationship listed in the roster for a particular roster member was not possible given the roster member's age and relationship to the respondent. In the majority of cases where the relationships could not be determined, this resulted in setting the relationship code to bad data and sometimes setting the roster member's age to bad data as well. In general, no effort was made to try to match the values of roster-derived household composition variables between pair members, because interviews of the different members of the same household could have taken place at different times. However, information from other pair members was sometimes used to change a relationship code from one value to another, instead of setting the relationship code to bad data.

11.2.2 Creating the Pair Relationship Variable (PAIRREL)

The process of identifying the pair relationships involved two steps: (1) match the household rosters of the pair members, and (2) determine the pair relationships using the relationship codes and ages of the matched rosters, if they could be determined. The first step is described in Section 11.2.2.1 and Appendix H, and the second step is described in Section 11.2.2.2 and Appendix H. Any relationships that could not be determined by this process were imputed as described in Section 11.5.2.

Because the creation of the multiplicity factors requires complicated programming logic, multiplicities could not be created for all possible pair relationships. The following pair relationships were considered to be of higher analytic interest, requiring the creation of multiplicities in each case:

- parent-child, child aged 12 to 14;
- parent-child, child aged 12 to 17;
- parent-child, child aged 15 to 17;
- parent-child, child aged 12 to 20;
- sibling-sibling, younger sibling aged 12 to 14, older sibling aged 15 to 17;
- sibling-sibling, younger sibling aged 12 to 17, older sibling aged 18 to 25;
- spouse-spouse (includes partner-partner), with children younger than 18;[142] and
- spouse-spouse (includes partner-partner), with or without children.

[142] The spouse-spouse pair relationship included respondents who were legally married, as well as respondents who lived together as though married (partners). Although the questionnaire distinguished between "spouses" and "partners," the pair relationship variable being described here did not distinguish between the two. In rare instances, a spouse-spouse pair included one pair member who identified the second pair member as a spouse, whereas the second pair member identified the first as a partner.

Even though these pair relationships were of the most interest, no restrictions were placed on the types of pairs that could be selected for inclusion in the NSDUH sample. However, the identification of the particular relationships between a given pair was limited by the relationship codes that were available: parent, child, grandparent, grandchild, sibling, spouse, unmarried partner, roommate, parent-in-law, child-in-law, boarder, other relative, and other nonrelative. (This precluded the possibility of identifying an uncle-nephew relationship, for example.) The various pair relationships that could be identified are stored in the variable PAIRREL, the levels of which are summarized in Table 11.3. The levels in PAIRREL do not correspond exactly with those given above, but the relevant pair relationships can be derived from the value of PAIRREL. For example, a value of PAIRREL = 3 indicates that, among the pair relationships given above, the pair relationship was a parent-child pair with a child aged 18 to 20.

Table 11.3 Levels of the Variable PAIRREL

Value of PAIRREL	Interpretation	Domain of Interest
1	Respondent is part of a parent-child (12–14) pair	Yes
2	Respondent is part of a parent-child (15–17) pair	Yes
3	Respondent is part of a parent-child (18–20) pair	Yes, indirectly
4	Respondent is part of a parent-child (21+) pair	No
5	Respondent is part of a sibling (12–14)-sibling (15–17) pair	Yes
6	Respondent is part of a sibling (12–17)-sibling (18–25) pair	Yes
7	Respondent is part of another sibling-sibling pair	No
8	Respondent is part of a spouse-spouse[1] pair, with children in the household younger than 18	Yes
9	Respondent is part of a spouse-spouse pair, with no children in the household younger than 18	Yes
10	Respondent is part of a spouse-spouse pair, but it is unclear whether children younger than 18 in the household belong to the pair	Yes
11	Respondent is part of a grandparent-grandchild pair	No
12	Respondent is part of another clearly identifiable pair	No
13	Respondent is part of a pair that is not clearly identifiable, but it is clear from the relationship codes that it is not within codes 1 through 11	No
14	Respondent is part of a pair that is not clearly identifiable, and it could be any pair relationship	Maybe
15	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 1 or 12	Maybe
16	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 2 or 12	Maybe
17	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 3 or 12	Maybe
18	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 4 or 12	No
19	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 5 or 12	Maybe

Table 11.3 Levels of the Variable PAIRREL (continued)

Value of PAIRREL	Interpretation	Domain of Interest
20	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 6 or 12	Maybe
21	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 7 or 12	No
22	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8 or 12	Maybe
23	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 9 or 12	Maybe
24	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 8, 9, or 12	Maybe
25	Respondent is part of a pair that is not clearly identifiable, but it could be pair codes 11 or 12	No
99	Respondent is not a member of a pair	No

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

11.2.2.1 Matching the Household Rosters

For the purpose of discussing how to match the household rosters in this report, the pair members are identified as pair member "A" and pair member "B." For the household roster of pair member A, it was necessary to determine which household member listed in A's roster corresponded to the other selected pair member. The same had to be done for pair member B. This was accomplished using the age and gender of the pair members, in addition to the variable MBRSEL, which was used to identify the roster member corresponding to the other selected pair member.

In a perfect setting, the questionnaire age and gender of pair member B (AGE and IRSEX, respectively) would have corresponded exactly to the age and gender entered for one of the members of pair member A's household roster (RAGE and RSEX). Moreover, the value of MBRSEL for this matched roster member would have been 1, and the value of MBRSEL for all other roster members would have been zero or missing. Furthermore, in this perfect setting, matches with exactly one MBRSEL = 1 correctly identifying the other pair member also would have been found with pair member B's roster. This did not always occur, however, so some effort was required to determine the roster member most likely to correspond to the other selected pair member.

The quality of the match between pair members varied depending upon the quality of the roster entries and the time between interviews. A number of if-then-else conditions, called priority conditions (because of the hierarchical nature of the conditions), gave a pair match that was considered valid in the vast majority of cases. These conditions are provided in Appendix H. In general, the conditions matched IRSEX and AGE for the one pair member against the age and gender of the roster members in the other pair member's roster, using MBRSEL to help identify

the appropriate roster member. For a match to be considered valid, it was necessary that at least one of the two pair members have a match.

In the cases where a single roster member had to be selected among duplicates and where the duplicates had the same relationship code, it was necessary to limit the relationship codes to child or sibling. In some cases, because of the poor quality of the rosters of the pair members, it was not possible to locate the household member listed in A's roster that corresponded to pair member B, and vice versa. The determination of the pair relationships for these cases was left to imputation. Even when a pair of roster members was successfully identified, it was not always possible to successfully determine the pair relationship, as described in the next section.

11.2.2.2 Determining the Pair Relationship Using the Relationship Codes of the Matched Rosters

Once the pair was identified, two observations per household resulted, each with a relationship code corresponding to the other selected pair member. The relationship codes for these two observations had to be matched to determine the pair relationship. For example, suppose a 15-year-old and a 38-year-old were selected to be interviewed. If the 38-year-old was subsequently identified as the parent on the 15-year-old's roster, and the 15-year-old was identified as the child of the 38-year-old on the 38-year-old's roster, then the pair relationship would be identified as PAIRREL = 2, according to the levels of PAIRREL provided in Table 11.3. Thus, these two individuals would belong to the following pair relationships of interest: parent-child with child aged 15 to 17, parent-child with child aged 12 to 17, and parent-child with child aged 12 to 20. As noted earlier, the pair relationship of interest was derived from the values of PAIRREL. In particular, the parent-child with child aged 12 to 17 and parent-child with child aged 12 to 20 domains were derived from the parent-child pair relationships created using 12- to 14-year-olds, 15- to 17-year-olds, and 18- to 20-year-olds, the levels referenced in PAIRREL. Moreover, the overall spouse-spouse domain was derived from the two spouse-spouse pair relationships with and without children.[143]

As with the procedure used to match the household rosters, a series of conditions was used to identify the relationship between pair members. These priority conditions used ages and relationship codes to identify the pair relationships. In a perfect setting, the relationship codes would be nonmissing and in agreement between the pair members, as in the example given in the previous paragraph. In some instances, however, either the relationship codes were missing or they did not agree across the pair members. The priority conditions offer a method for interpreting the relationship codes in such cases.

Below are the strategies used to identify a pair relationship in an imperfect setting:

1. If a relationship code was missing on one side of the pair but not on the other, then the pair relationship was assumed to be identified by the nonmissing relationship code. The exception to this rule occurred if the identified relationship was parent-child with a child younger than 18, the "parent" was less than 10 years older than the child, and the "parent" answered the parenting experiences question (FIPE3) by

[143] The spouse-spouse pair relationship includes partner-partner pair relationships.

saying that the other respondent was not his or her child. In this case, the nonmissing relationship code was considered spurious, and the relationship was left missing.

2. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but the relationship codes on both sides indicated that the unknown pair relationship was not a relationship of interest, then the pair relationship was identified as such and no imputation was required. For example, if pair member "A" identified pair member "B" as a "boarder," but pair member "B" identified pair member "A" as "other relative," then the relationship was not a relationship of interest and the variable PAIRREL would have been assigned the value "13" (Table 11.3).
3. If it was not possible to definitively determine the relationship between the pair members using the relationship codes, but a parent-child relationship was possible given the relationship code in one of the pair member's rosters, then the FIPE3 variables were used to assist in the determination of a pair relationship. For example, consider a pair member who was a stepparent and refers to his or her stepchild as "child," but the child refers to the stepparent as "other nonrelative." Membership in a parent-child relationship where the child was younger than 18 was indicated if the stepparent answered FIPE3 affirmatively, thereby proceeding to the parenting experiences module. On the other hand, if the stepparent answered FIPE3 negatively, then the stepparent was not considered the parent. A third scenario arose if the FIPE3 answer was not given. In this case, a parent-child relationship was assumed if the stepparent was legally married and the child identified the spouse of the other pair member as "parent."

The quality of the match for PAIRREL levels 1 through 25 is indicated by the variable RELMATCH, the levels of which are summarized in Table 11.4. In general, imputation was required for values of RELMATCH of 0 or 4, or if PAIRREL = 10. PAIRREL = 10 was a special case because it was clear that a relationship "of interest" always would have been involved. For this value of PAIRREL, the value of RELMATCH was equal to 1 or 2. However, imputation was still required because it was not clear whether children were in the household. The number of cases that were matched or not matched, as indicated by the RELMATCH variable (or PAIRREL = 10), for the 2012 survey is provided in Table 11.5. The amount of imputation required was dependent upon the quality of the rosters. The attributes of the roster are described in Chapter 8.

Table 11.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH

Value of RELMATCH	Values of PAIRREL	Interpretation
0	14	FAILURE: The relationship was not identifiable and could have been a relationship of interest.
1	1–9, 11–13	SUCCESS: The relationship was clearly identifiable using information from both pair members or was unmistakably not a relationship of interest.
1	10	FAILURE: A spouse-spouse[1] relationship was definitively established using information from both pair members, but it was unclear whether the pair had children in the household.

Table 11.4 Values of PAIRREL That Correspond to the Levels of the Variable RELMATCH (continued)

Value of RELMATCH	Values of PAIRREL	Interpretation
1.5	8	SUCCESS: A spouse-spouse relationship was definitively established using information from both pair members, and children younger than 18 were in both rosters. Relationship codes on one side of the pair indicated children belonged to the pair, and on the other side of the pair, the relationship codes corresponding to the children were missing.
2	1–9, 11–13	SUCCESS: The relationship was clearly identifiable using information from one pair member, while the relationship code from the other pair member was missing.
2	10	FAILURE: A spouse-spouse relationship was definitively established using information from one pair member, while the relationship code from the other pair member was missing. It was unclear whether the pair had children in the household.
3	1–8, 12, 13	SUCCESS: Relationship information was conflicting between the pair members, but conclusions were drawn anyway for some parent-child pairs, some sibling-sibling pairs, and some spouse-spouse pairs using either information outside the household roster or logical reasoning.[2]
4	15–25	FAILURE: Relationship information was not identifiable. Information was in conflict between the pair members, where one pair member indicated relationship of interest and the other did not. However, ages supported a relationship of interest (may be used to limit imputation).

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships.

[2] In the case of potential parent-child pairs, further evidence that a parent-child relationship was involved or not involved was obtained by looking at the FIPE3 variable to see whether a stepparent had a spouse who corresponded to a child's parent or to see the ages of the respondents. For spouse-spouse relationships, two situations occurred. In the case where the respondents were not legally married, the children of one pair member were considered the children of the pair in the household, even though they were not identified as such by the other pair member. In the case where only one pair member referred to the other as a "married" or "unmarried partner," and if both had the same children, they were considered "spouse-spouse-with-children." The other pair member was usually referred to as a "roommate" or "other nonrelative."

Table 11.5 Frequencies of the Levels of the Variable RELMATCH: 2012

RELMATCH	Frequency (Percent)	Requires Imputation
0	25 (0.13)	Yes
1 (PAIRREL ≠ 10)	19,062 (97.96)	No
1 (PAIRREL = 10)	19 (0.10)	Yes
1.5	0 (0.00)	No
2 (PAIRREL ≠ 10)	91 (0.47)	No
2 (PAIRREL = 10)	2 (0.01)	Yes
3	87 (0.45)	No
4	173 (0.89)	Yes

11.3 Stage Two Editing: Multiplicity Counts

As stated earlier, multiplicities were required to account for analyses that were performed at the person level, even though the pair weights were calculated at the pair level. Because the multiplicities were relevant only at the person level, the definition of multiplicity required the identification of the focus member of the pair. Using the pair relationships determined in Section 11.2, the following domains were considered:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 15 to 17), parent focus;
4. parent-child (child 15 to 17), child focus;
5. parent-child (child 12 to 17), parent focus;
6. parent-child (child 12 to 17), child focus;
7. parent-child (child 12 to 20), parent focus;
8. parent-child (child 12 to 20), child focus;
9. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
10. sibling (12 to 14)-sibling (15 to 17), sibling (12 to 14) focus;
11. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
12. sibling (12 to 17)-sibling (18 to 25), sibling (12 to 17) focus;
13. spouse-spouse (includes partner-partner) with children younger than 18; and
14. spouse-spouse (includes partner-partner).

Determining the multiplicity entailed finding the number of roster pairs in the domain of interest that contained the focus member in the pair. In broad terms, the process of determining the multiplicity count involved two steps: (1) calculate the multiplicity count for each pair member, and (2) use the screener, quality of roster, and other means to determine the appropriate count if each pair member's counts did not match. The first step is described in Section 11.3.1, and the second step is described in Section 11.3.2 and Appendix I. Multiplicities that could not be determined through these steps were left to imputation as described in Section 11.5.3.

Because the pair weights reflect selection done at the time of screening, the multiplicity count should reflect the household makeup at that time. However, this was not possible in all cases, because for some households the screener roster was not as complete as the questionnaire roster, and recorded relationships in the screener roster were relative to the head of the household rather than to each pair member. No account was made for cases where a change in the household makeup occurred between the screening date and the dates of both interviews. In other words, due to the passage of time only, the observed change in household makeup could have occurred because of an intervening birthday or because a roster member left or entered the household after screening. Because an adjustment to the multiplicity counts would have been extremely complicated to implement for the small number of cases to which it applied, no such

adjustment was made. Nevertheless, when possible, the screener was used to resolve cases where there were disagreements between pair members on the value of the multiplicity count.

11.3.1 Determining the Multiplicity Count for Each Pair Member

The multiplicity counts for each pair member consisted of a direct count and an indirect count. The direct count was obtained by looking at the focus pair member. It was a count of the roster members who could have been selected, where the same pair domain would have resulted. The indirect count was obtained by looking at the pair member who was not the focus. It was a count of the pair member himself or herself, plus other roster members who, by virtue of their relationship code, would have had the same pair relationship had they been selected.

A summary of the ways of determining the direct count and indirect count for each pair domain are provided in Table 11.6. For these domains, neither the direct nor the indirect count could be zero, because the pair member who was not the focus had to be part of the count. For spouse-spouse counts, no work was necessary to determine multiplicity counts. If a respondent was in a spouse-spouse pair, the multiplicity count was necessarily 1 in almost all cases, because only one spouse-spouse pair could have been selected that included that pair member. If the true multiplicity count exceeded 1, then the multiplicity count was set to 1.[144] Note that other spouse-spouse pairs in the household (e.g., one spouse's parents) would have been of interest in the household counts discussed in subsequent sections.

Table 11.6 Multiplicity Counts for Each Pair Member

Pair Relationship	Focus Member	Direct Count	Indirect Count
Parent-Child	Child	From child: number of parents	From parent: self + spouse/partner
Parent-Child	Parent	From parent: number of children in appropriate age range	From child: self + number of siblings in the appropriate age range
Sibling-Sibling	Older sibling	From older sibling: number of siblings in younger age range	From younger sibling: self + number of siblings in younger age range
Sibling-Sibling	Younger sibling	From younger sibling: number of siblings in older age range	From older sibling: self + number of siblings in older age range

11.3.2 Determining the Final Multiplicity Count

Once the counts were determined for each pair member, it was necessary to resolve any differences between these counts across pair members. In most cases, the direct and indirect counts agreed, with no bad relationship codes for either pair member, resulting in a straightforward determination of the final multiplicity count. This was usually possible if one pair member had bad relationship codes or had a count of zero, which meant that the final

[144] In rare cases, it was possible for a respondent to have two or more spouses who lived in the same household. Determining the appropriate multiplicity count in these cases required knowledge of which spouse was the focus, which would be arbitrary. Because having multiple spouses was an extremely rare occurrence, and because of the complexity of determining the appropriate multiplicity count, these situations were not explicitly addressed during data processing.

multiplicity count came from the pair member with good data.[145] For some cases, both pair members had bad relationship codes, which meant that the final multiplicity was left to imputation. Among the remaining cases, some could be reconciled and some could not. In the cases where reconciliation was possible, many of the disagreements between the pair members were resolved by going to the screener. The method used to reconcile differing counts depended upon the domain. For the parent-child domains, for example, in addition to the screener, the parent/legal guardian variable, FIPE3, was used to help reconcile differences. Detailed rules for reconciling differences between pair members are provided in Appendix I.

If reconciliation between the counts from the two pair members in the household and the screener was not possible, upper and lower bounds within which the imputed value had to reside were determined from the counts for each pair member and the counts for the screener. The amount of imputation required for the multiplicity counts for the 2012 survey is shown in Table 11.7. From this table, it is apparent that the greatest degree of uncertainty came with the determination of the number of parents in the child-focus parent-child domains. This occurred because, even though the parent-child pair relationship had been established, it often was unclear whether there was a second "parent" in the household.

Other domains had very little uncertainty. The counts of the number of children in the parent-focus parent-child domain, for example, were almost always definitively determined.

Table 11.7 Amount of Imputation Required for Multiplicities in Various Pair Domains: 2012

Pair Domain	Multiplicity	Missing Cases
Parent-Child (12–14), Child Focus	Number of parents	144
Parent-Child (12–14), Parent Focus	Number of children	50
Parent-Child (15–17), Child Focus	Number of parents	134
Parent-Child (15–17), Parent Focus	Number of children	50
Parent-Child (12–17), Child Focus	Number of parents	228
Parent-Child (12–17), Parent Focus	Number of children	50
Parent-Child (12–20), Child Focus	Number of parents	60
Parent-Child (12–20), Parent Focus	Number of children	60
Sibling (12–14)-Sibling (15–17), Older Sibling Focus	Number of younger siblings	52
Sibling (12–14)-Sibling (15–17), Younger Sibling Focus	Number of older siblings	52
Sibling (12–17)-Sibling (18–25), Older Sibling Focus	Number of younger siblings	58
Sibling (12–17)-Sibling (18–25), Younger Sibling Focus	Number of older siblings	58

11.4 Stage Three Editing: Household-Level Person Counts

In order to improve the quality of the estimates from the pair data through poststratification of the appropriate weights (Westlake, Chen, & Gordek, 2014), it was necessary to identify the household-level person counts for each domain. This entailed finding the number

[145] There were some exceptions to this rule. If the bad relationship codes were only within the relevant age ranges, then the count from the good side was used only if the age ranges in the good side matched the screener.

of individuals in the household who belonged to a particular domain, given that one member of a domain was selected as the focus. These counts were more difficult to derive than the multiplicity counts because all households were considered. Within each household, counts for any of the domains of interest were derived, regardless of whether the pair belonged to that domain or even whether a pair was selected at all. The counts were derived for 10 of the 14 pair domains described in Section 11.3. For the parent-child counts where the child was between 15 and 17, calculating the household counts was unnecessary.[146] For the other two remaining sibling-sibling domains, the reason is historical: They were added after the procedures were first developed, and there was insufficient time to develop the household counts for those domains. The domains where these counts were created are listed below:

1. parent-child (child 12 to 14), parent focus;
2. parent-child (child 12 to 14), child focus;
3. parent-child (child 12 to 17), parent focus;
4. parent-child (child 12 to 17), child focus;
5. parent-child (child 12 to 20), parent focus;
6. parent-child (child 12 to 20), child focus;
7. sibling (12 to 14)-sibling (15 to 17), sibling (15 to 17) focus;
8. sibling (12 to 17)-sibling (18 to 25), sibling (18 to 25) focus;
9. spouse-spouse (includes partner-partner) with children younger than 18; and
10. spouse-spouse (includes partner-partner).

Determining the household-level person counts was a two-step process: (1) calculate the household count for each respondent, whether a member of a pair or a single respondent; and (2) use the screener, quality of roster, and other means to determine the appropriate final count either by attempting to reconcile differing counts between pair members or by attempting to determine the appropriate count when information from only one roster was available. For households where only one respondent was selected, the matching step (step 2) was unnecessary. The first step is described in Section 11.4.1, and the second step is described in Section 11.4.2 and Appendix J. Household counts that could not be determined by this process were left up to imputation as described in Section 11.5.4.

Because the pair weights reflected selection at the time of screening, the household-level person counts should have reflected the household makeup at that time. As with the multiplicity counts, however, this was not entirely possible, so no account was made for cases where a change in the household makeup occurred between the screening time and the time of each interview. An explanation for why this was not possible for the multiplicity counts is described in Section 11.3. Moreover, as stated in that section, to implement such an adjustment would have

[146] Because household counts were defined for everybody, it was possible to derive these counts using the counts for the parent-child domains where the child was between 12 and 14 and where the child was between 12 and 17. However, the multiplicity counts for the parent-child (15 to 17) domain had to be calculated and could not have been derived in as straightforward a way. This was because the multiplicity counts were only defined if the pair relationship corresponded to the pair domain of interest.

been extremely complicated for the household-level person counts. Nevertheless, in cases where there were disagreements between pair members on the value of the household-level person count, the screener was used to resolve those disagreements.

11.4.1 Determining the Household-Level Person Count for Each Respondent

11.4.1.1 Parent-Child Domains

When obtaining household-level person counts for parent-child domains, the six parent-child domains previously listed were considered. In any household, the household-level person counts for parent-child domains were nonzero if at least one parent was present in the household with children within the relevant age range. In this instance, the child-focus count would have been the number of children in the household within that age range that belonged to the parent in the household, and the parent-focus counts would have been the number of parents. If more than one "family unit" (mother and/or father with children) lived within the household, the child-focus counts should have counted children from more than one set of parents, and the parent-focus counts should have counted two or more parents, at least one for each set of children.

One situation where this could occur was where three generations lived within the same household, with children in both the youngest and the second generations within the relevant age range. Using the youngest generation as the reference point, some of the parent's siblings (the grandparents' other children) were within the relevant age range. In this instance, the parent-child domains of the number of children would have included both the children of the parents and the children of the grandparents who were in that age range. The count of the number of parents included both the parents and the grandparents (and exceeded 2). Identifying more than one family unit in a household with children within the relevant age range under other scenarios (e.g., two sisters both with children within the relevant age range, both living within the same household) could not be determined from the data and had to be disregarded. Regardless of how many family units were in the household, counts had to be determined in different ways depending upon whether a parent-child pair "of interest" was selected or not.

Descriptions of how to obtain the household-level person counts are provided below for the parent-child domains outlined above. Parent-child pairs of interest with parent-focus and child-focus domains considered together are described first. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This is followed by descriptions for other pairs and single respondents with parent-focus and child-focus domains considered separately.

11.4.1.1.1 Obtaining Counts for Parent-Child Domains (Parent-Focus and Child-Focus): Parent-Child Pairs, Child Younger than 21

If the pair was identified as parent-child and the three-generation situation described above was not apparent, the household-level child-focus person count was given by the parent-focus multiplicity count. Similarly, the household-level parent-focus person count was given by the child-focus multiplicity count. If a three-generation situation was identified and the grandparent also had children within the relevant age range, the number of children and the number of parents were adjusted appropriately. The final household count in this instance was

greater than the imputation-revised multiplicity count, which did not include all of the children in the household within the relevant age range.

11.4.1.1.2 Obtaining Counts for Child-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs[147] and single respondents, the following conditions were required to determine the household count for the number of children of parents in the household:

1. If the age of the respondent was within the relevant age range and that respondent had at least one parent, then the child-focus counts were determined in the same way as the parent-focus multiplicity counts: The count was of the "self" plus the respondent's siblings within the relevant age range. If the respondent's parents were not identified as living with him or her in the household, then the count was set to zero.
2. If the respondent had children within the relevant age range, then the count was of the respondent's children within that range. If the respondent also had older children who had children of their own within the relevant age range, then the count was of the respondent's children and grandchildren within the relevant age range.
3. If the age of the respondent was outside the relevant age range, but the respondent had parents living with them in the household and had siblings within the relevant age range, then the count was of the number of the respondent's siblings.
4. If the respondent had grandchildren within the relevant age range and the respondent also had children older than 25 or children-in-law living with them, then the count was the number of the respondent's grandchildren. The assumption was that the respondent's children or children-in-law were the parents of the respondent's grandchildren. The likelihood of this not being the case was small. In the case where a pair was selected, this was resolved by looking at the count of the other pair member.

11.4.1.1.3 Obtaining Counts for Parent-Focus Parent-Child Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count for the number of parents of children in the household:

1. If the age of the respondent was within the relevant age range, then the count was of the number of the respondent's parents (which could be zero).
2. If the age of the respondent was outside the relevant age range but the respondent had siblings within the relevant age range, then the count was of the number of the respondent's parents (again, this could be zero).

---

[147] "Other pairs" included pairs that were not within a domain of interest because the age of at least one of the pair members was outside the relevant age range. For parent-child pairs, this applied to a pair with a child who was 21 or older. For sibling-sibling pairs, this applied to siblings where both were within the same age range (both were 12 to 14, 15 to 17, or 18 to 25) or at least one of the siblings was older than 25. "Other pairs" also are referenced in Sections 11.4.1.1.3 and 11.4.1.2.2.

3. If the respondent had children within the relevant age range, then the parent-focus counts were determined in the same way as the child-focus multiplicity counts: The count was of the self plus the spouse or unmarried partner. If the respondent also had older children (older than 25 and living with him or her) who had children of their own (identified as grandchildren) within the relevant age range, then the count was at least two. If the respondent had a spouse or unmarried partner in the household, then the count was incremented by one, and if a child-in-law was in the household, then the count also was incremented by one. (Note that, under these scenarios, the number of parents could range between two and four.)
4. If the respondent had grandchildren within the relevant age range but no children in that range, and the respondent had a child older than 25 or a child-in-law living with them, then the count was 2 if both the child older than 25 and the child-in-law were living in the household, and the count was 1 if not.

11.4.1.2 Sibling-Sibling Domains

When obtaining household-level person counts for sibling-sibling domains, only the two sibling-sibling domains previously listed were considered. As with the parent-child counts, the household-level person counts for sibling-sibling domains were nonzero if at least one sibling-sibling pair was present in the household within the relevant age ranges, in which the count was the number of appropriately aged siblings. If sets of siblings from more than one "family unit" (sets of siblings from different parents) resided within the same household, the sibling-sibling counts should have counted possible pairs from within each set. However, sets of siblings that did not involve the respondent's family unit could not have been identified from the data. Regardless of how many sets of siblings were in the household, counts had to be determined in different ways depending upon whether a sibling-sibling pair "of interest" was selected or not.

Descriptions of how to obtain the household-level person counts are provided below for the sibling-sibling domains outlined above. Sibling-sibling pairs of interest are described first. In this instance, the pair actually belonged to a pair relationship where analysis using one or more of the domains listed was possible. This is followed by descriptions for other pairs and single respondents. In each case, the descriptions apply regardless of which sibling-sibling domain was considered.

11.4.1.2.1 Obtaining Counts for Sibling-Sibling Domains: Sibling-Sibling Pairs of Interest

If the pair was identified as sibling-sibling within a relevant domain, the multiplicity count was the number of younger siblings because the older sibling was the focus. The household-level sibling-sibling person counts were determined in a similar manner to the multiplicity count, except that the count of interest was for the number of older siblings. If the pair member was the older sibling, then the household count was the self plus the number of siblings in the older age range. The count for the younger sibling pair member was the number of siblings within the same older age range. Unlike the case with the parent-child household-level counts, inconsistencies in the sibling-sibling counts when the pair selected was sibling-sibling still needed to be resolved. However, the rules for resolving inconsistencies followed directly from those used for the multiplicity counts when counting the number of younger siblings

(Appendix I). Note that a pair that was within one sibling-sibling pair domain had to be outside the other sibling-sibling pair domain.

11.4.1.2.2 Obtaining Counts for Sibling-Sibling Domains: Other Pairs and Single Respondents

For other pairs and single respondents, the following conditions were required to determine the household count for the number of siblings within the older age ranges of the domains of interest in the household:

1. If the age of the respondent was within the age range of the older sibling and that respondent had at least one sibling in the younger age range, then the count was the self plus the respondent's siblings within the older age range. If the respondent did not have any siblings within the younger age range, then the count was set to zero.
2. If the age of the respondent was within the age range of the younger sibling and that respondent had at least one sibling in the older age range, then the count was the number of the respondent's siblings in the older age range.
3. If the age of the respondent was outside the age range of the older or younger sibling but had at least one sibling in each of the older and younger age ranges, then the count was the number of siblings in the older age range.
4. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had children within both the older and the younger age ranges, then the count was set to the number of respondent's children in the older age range.
5. If the age of the respondent was outside the age range of the older or younger sibling but the respondent had grandchildren within both the older and younger age ranges, then the count was the number of grandchildren in the older age range. If the respondent's grandchildren were cousins rather than siblings, then there was no way of deciphering this from the data. This had to be resolved by looking at the information from the other pair member, if another pair member was selected.

11.4.1.3 Spouse-Spouse Domains

What is referred to as a "spouse-spouse domain" was actually derived from spouse-spouse and partner-partner pair relationships. The following conditions were required for the number of spouse-spouse (including partner-partner) pairs to be incremented by 1. Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair.
2. The respondent was not part of a spouse-spouse pair but had a spouse (or unmarried partner).
3. The respondent had two parents living in the house.
4. The respondent had two parents-in-law living in the house.

5. The respondent had two grandparents living in the house.
6. The respondent had a child and a child-in-law living in the house.

The following conditions were required for the number of spouse-spouse pairs with children younger than 18 to be incremented by one. (These also include partner-partner pairs with children younger than 18.) Some of these conditions were applied to the same household:

1. The respondent was part of a spouse-spouse (or partner-partner) pair with children younger than 18.
2. The respondent was not part of a spouse-spouse pair[148] but had a spouse (or unmarried partner) and children younger than 18.
3. The respondent had two parents living in the house and was either younger than 18 or had siblings younger than 18.
4. The respondent had a child and a child-in-law living in the house and had grandchildren younger than 18.

11.4.2 Determining the Final Household-Level Person Count

For a particular type of household-level person count, there are three types of households from a sample selection perspective. For the first type, a pair was selected where the pair relationship corresponded directly to the pair domain being counted and both pair members responded. In this case, the household-level person count was usually easy to obtain using the multiplicity counts, although an adjustment was sometimes required if more than one family unit was in the household. For example, if a parent-child pair was selected where the child was 12 years old, the household-level person counts for the parent-focus parent-child (12-14) domain could usually be obtained from the multiplicity count that was calculated in stage two. In the second type of household, a pair also was selected and both pair members responded, but in this type the pair relationship did not correspond directly to the pair domain being counted. In this case, determining the final count was sometimes more difficult, particularly if one or more of the counts was a count of zero. A count of zero from a roster with good data did not necessarily mean that the final count should be zero. For example, suppose a household consisted of a man, his wife, brother, and two sons, and suppose one of the sons and his uncle (the man's brother) were selected. If the uncle's roster had a count of zero for all domains of interest—because all of the household members were "other relatives" from his perspective—then no nonzero parent-child count could be obtained. The final count would have to be determined from imputation. In the third type of household, only one respondent was selected. In this case, it was not possible to match counts from different pair members, but determining the final count could still be difficult if the count was zero for a household where the value was not truly zero.

For situations where a pair was selected and both pair members had good roster data, if the counts agreed between the pair members and were not zero, then a straightforward determination of the final household-level count was possible. This occurred in a majority of

[148] All spouse-spouse pairs were excluded here because spouse-spouse pairs with children were already accounted for, and spouse-spouse pairs without children had already been defined (possibly by imputation) not to have children younger than 18.

cases. If one pair member had a bad roster with no information in it and the other had a good roster, this was treated in the same way as if a single respondent was selected with a good roster. In either of these cases, the final count could be determined, provided a considerable number of conditions were satisfied. The conditions used to accept a good roster's count, when either the other pair member's roster was bad or no pair was selected, are provided in Appendix J. If these conditions were not met, the final household-level person count was left to imputation. Imputation also was required if two pair members were selected, both with bad rosters.

Among the remaining cases, some could be reconciled and some could not. In the cases where reconciliation was possible, some of the disagreements were caused by the pair members' rosters having different age and gender compositions. In these cases, many of the disagreements between the pair members were resolved by going to the screener. However, the screener did not provide much help if the age and gender composition of the pair members' rosters were identical, yet the counts still disagreed, as was the case with the uncle-nephew pair described above. In that example, one count was zero and the other was nonzero. Under conditions set out in Appendix J, it was possible to determine that the disagreement in this case was due to the uncle not being able to identify the parent-child domains, and the nonzero count was used. More detailed rules for reconciling differences between pair members are described in Appendix J.

If the attempt to reconcile differences in the household-level person counts between pair members was unsuccessful, upper and lower bounds within which the imputed value must reside were determined from the counts for each pair member and the counts for the screener.

11.5 Imputation for the Pair Variables

Imputation was required for variables in all three stages, and the imputation models for the pair variables required the inclusion of covariates at the pair level. The creation of these covariates is described in Section 11.5.1. The imputation process for stages one, two, and three are described in Sections 11.5.2 through 11.5.4, respectively. The final covariates for the fitted models are listed in Appendix D.

11.5.1 Creation of Covariates for Imputing Pair-Level Variables

Imputation was performed at the household level rather than at the respondent level. Thus, it was necessary to model covariates defined at the household level. Segment-level covariates were used for this purpose because they were automatically defined at the household level, using external information that was constant regardless of when the interviews were conducted. In addition to these segment-level covariates, information from the questionnaire would also have been useful as modeling variables. The logical choices for questionnaire-derived variables include the household composition variables IRHHSIZE (household size), IRKID17 (number in household younger than 18), IRHH65 (number in household aged 65 or older), and IRFAMSKP (presence of other family members in household indicator).

However, because interviews between pair members could have been conducted at different times, these variables were not necessarily consistent across pair members. Therefore, new count variables were needed that were consistent across the pair members (i.e., used screener information to reconcile disagreements between them) within a household. These

variables were created in a two-step process: (1) create the count variables for each pair member, and (2) attempt to reconcile disagreeing values between pair members. The following sections describe these two steps in the creation of household size, household composition age count variables, and household composition age count variables "for males only," each of which were consistent across pair members. Note that household composition age count variables were not created for females because, for a given age range, the number of females could be obtained by subtracting the number of males from the total number within that age range. These variables also had to be created for respondents who were not part of a pair, for the purposes of creating and imputing the household-consistent person counts of various domains.

11.5.1.1 Household Size

The new variable created to represent a household size that was consistent across selected pair members was called HHSIZE and constructed as follows. First, the edited household size, TOTPEOP, was compared between pair members. If the values for TOTPEOP agreed across pair members and were both nonmissing and greater than 1,[149] then HHSIZE was set to that value. If the values of TOTPEOP disagreed across pair members because the count for one pair member was missing and the count for the other was not missing and was greater than 1, then a natural choice for HHSIZE was the valid, nonmissing value contributed by one of the pair members. If the values of TOTPEOP either (1) disagreed across pair members without a clear indication of which one was valid or (2) were both missing or equal to 1, then the tools used to determine the final value of HHSIZE included the reported and edited household size variables, QD54 and TOTPEOP, as well as other measures of household size and "quality of roster" measures. These "other measures" included the screener household size and two sums of total valid ages within a pair member's roster.

The first sum was a simple total count of the number of roster members with valid ages, obtained by summing the counts within age groups. The second sum counted the ages of the pair members as reported during questionnaire administration. The two sums differed if a roster count (the first sum) was less than the number of pair members in a given age category. For example, if a household roster had one 12- to 17-year-old, but two 12- to 17-year-olds were selected, then the value of HHSIZE would be increased by 1. An additional situation occurred where the household size counts could not be easily determined by looking at both pair members. If the counts for both pair members were missing, then the screener household size was used to define HHSIZE.

In still other cases, disagreement between pair members with regard to the true household size could not be easily resolved. The screener household size did not support either household size in these cases, and the age counts mentioned above also did not resolve the disagreement. A decision had to be made as to which pair member's household size should be believed. This decision depended upon the "quality of the roster," where the household size was determined by the pair member with a "better" roster quality. One obvious way to measure roster quality was by noting the number of cases where the ages, relationship codes, or genders were missing in the roster. If a roster was missing one or more of these three variables for some of the roster

[149] In households where a pair of respondents was selected, the household size had to be greater than or equal to two.

members, the roster was considered to be of "poorer quality" than a roster with these variables nonmissing for all roster members.

If only one household member was selected as a respondent, referred to as a "nonpair household," the rules for creating HHSIZE were the same as those that were used if two household members were selected in a pair but only one of the pair members had a nonmissing, acceptable value for a reported household size. Note that if only one household member was selected as a respondent, it was permissible to have a reported household size of one, whereas in a selected pair, a reported household size of one was considered "bad data."

In summary, the variables used to determine HHSIZE included (for each pair member) the reported and edited household sizes, the number of cases with valid ages in the roster, the number of cases with valid ages with the count in some age categories replaced by the minimum possible in that age category, and a roster quality count of the number of roster members with missing information. The screener household size, which was the same for each pair member, also was used. Using all of these tools, HHSIZE did not have any missing values in the 2012 survey, nor did it have any in surveys from previous years. General points about the creation of the household size variable are provided in Appendix K.

11.5.1.2 Household Composition Age Count Variables

It would seem logical to assert that the ages of other household members would be good covariates for the domain to which a pair might belong. Such variables also would be important for imputing multiplicity and household-level domain counts. The household-consistent age counts were limited to the following age ranges: younger than 12, 12 to 14, 15 to 17, 12 to 17, 12 to 20, 18 to 25, 26 to 34, 35 to 49, and 50 or older. These variables were called AGE011, AGE1214, AGE1517, AGE1217, AGE1220, AGE1825, AGE2634, AGE3549, and AGE50P, respectively.

The first step in this process was to count the nonmissing ages for roster members in the household for each pair member. In some cases, it was necessary to adjust the count because the ages could not be matched exactly. For example, suppose a 38-year-old and a 17-year-old were interviewed, and the 17-year-old was interviewed first. Suppose also that the 17-year-old turned 18 (i.e., had his or her 18th birthday) before the 38-year-old was interviewed. Hence, the 17-year-old would have had an age of 18 in the 38-year-old's roster. However, because the younger pair member was 17 at the time of his or her interview, the ages of interest for this pair domain were defined to be 17 and 38. Hence, it was necessary to account for this by creating a new roster age variable that matched the age provided in the other pair member's questionnaire. The age counts using this new roster age variable were equivalent to subtracting 1 from the previously obtained 18- to-25 count and adding 1 to the previously obtained 12-to-17 count in the 38-year-old's roster. These adjustments were made for all cases where a match was made between one pair member's roster and another pair member's interview age and gender and the ages did not match exactly.

If no roster ages were missing, the sum of these counts was equal to the edited household size TOTPEOP. Note that the reported household size was not considered here, because the counts were obtained from an edited roster. As with household size, a series of priority

conditions was used to obtain the most likely count within each age group. If the appropriate count was ambiguous because of disagreement between the pair members, the quality of the roster and the age of the respondent (in that order) were used to determine the appropriate count. The roster quality was determined by the number of bad or missing roster entries and the quality of the match between the pair member's roster and the other pair member's questionnaire age and gender.

If only one household member was selected as a respondent, the rules were the same as when two household members were selected in a pair but only one of the pair members had nonmissing data for the roster ages. One important exception to these rules was that when determining minimum possible counts for various age groups, it was not necessary to incorporate information from another pair member to increment the minimum for that pair member. General points about the creation of the age variables are provided in Appendix K.

11.5.1.3 Household Composition Age Counts of Males

For some pair variables, particularly spouse-spouse pairs, knowledge of the gender of the roster member was important for imputation purposes. In a similar manner to that used in the creation of the household composition age count variables, variables counting the number of males within the given age ranges were created. Disagreements between pair members were resolved in a similar manner to what was done with the household composition age count variables. The names of the male age counts were MALE011, MALE1214, MALE1517, MALE1217, MALE1220, MALE1825, MALE2634, MALE3549, and MALE50P.

11.5.2 Stage One Imputation: Pair Relationships

Missing pair relationships were filled in using the single response propensity (RP)/single prediction (PRD) type of the PMN method. In this stage, the imputation set involved only the edited variable PAIRREL, with its corresponding imputation-revised variable IRPRREL. Because the pair relationship varies according to the ages of the respondents, modeling and imputation were done independently within each of 11 age group pairs. Table 11.8 presents these 11 age group pairs, as well as the pair relationships prevalent within each age group pair.

Table 11.8 Age Group Pairs with Associated Possible Pair Relationships

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1]	
		≥ 10% Prevalence[2]	< 10% Prevalence
0	12–14/12–14	Sibling-sibling	Other relationship
1	12–14/15–17	Sibling-sibling	Other relationship
2	12–14/18–25	Sibling-sibling	Other relationship; parent-child; spouse-spouse**
3	15–17/15–17	Sibling-sibling	Other relationship; spouse-spouse*
4	15–17/18–25	Sibling-sibling	Other relationship; spouse-spouse; parent-child*
5	18–20/18–25	Other relationship; sibling-sibling; spouse-spouse	Parent-child**

Table 11.8 Age Group Pairs with Associated Possible Pair Relationships (continued)

Age Group Pair Number	Age Group Pair	Pair Relationships Appearing in Age Group Pair (in Order of Prevalence)[1]	
		≥ 10% Prevalence[2]	< 10% Prevalence
6	21–25/21–25	Spouse-spouse; other relationship; sibling-sibling	Parent-child**
7	12–14/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling*
8	15–17/26+	Parent-child	Other relationship; grandparent-grandchild; sibling-sibling; spouse-spouse**
9	18–20/26+	Parent-child	Other relationship; sibling-sibling; spouse-spouse; grandparent-grandchild
10	21+/26+	Spouse-spouse; parent-child; other relationship; sibling-sibling	Grandparent-grandchild*

* Pair relationships occur in less than 1 percent of the overall total number of pair relationships.
**The pair relationship is so rare that it does not appear in the age group pair in every survey year.

[1] The pair relationship labeled "spouse-spouse" includes partner-partner pair relationships. The spouse-spouse domain as listed here actually consists of two domains (spouse-spouse-with-children and spouse-spouse-without-children) that have been collapsed for the purposes of making the table easier to read. "Other relationship" refers to a relationship other than sibling-sibling, parent-child, grandparent-grandchild, or spouse-spouse.

[2] The pair relationships each form at least 10 percent of the overall total number of pair relationships within the given age group pair, and the total is at least 85 percent of the overall total.

11.5.2.1 Response Propensity Step

For a respondent pair to be considered complete, the pair relationship must be definitively established. In terms of the variable PAIRREL, this meant that the pair had to have a value of PAIRREL within the range of 1 to 9 or equal to 11 or 12. A value of PAIRREL equal to 13 also was considered complete, even though the pair relationship was not definitively established, because it was known that the pair relationship was not a relationship of interest. Response propensity adjustments then were calculated for each age group pair in order to make the respondent pair weights representative of the entire sample of pairs. Because the modeling of the final pair weight adjustments was not completed at the time of the pair imputations, the pair-level sample design weights were adjusted to account for nonresponse at the household level using a simple ratio adjustment. These adjustments were calculated using an item response propensity model, which is a special case of the generalized exponential model. See Appendix A in Westlake et al. (2014) for technical details about this procedure.

11.5.2.2 Prediction Step

After the weights were adjusted using the item response propensity model within each age group pair, logistic regression models were fitted using the adjusted weights. Preferred covariates in these models included the age count variables described in Section 11.5.1.2. However, variables with missing values for some observations cannot be used as covariates in the models. To allow for better models whenever possible, the data were partitioned into two

groups: those with nonmissing values for all age count variables, and those where one or more values of the age count variables were missing. When all of the age count variables were nonmissing, the predicted mean model was fitted using these variables. Otherwise, the model was fitted using the overall household count. This resulted in two predicted mean models for each of the 11 age group pairs.[150] These groups were combined again before the hot-deck imputation step. Further details about the variables used in these models can be found in Appendix D.

All modeling was done using the SUDAAN procedure MULTILOG; however, the number of levels in the response variable varied based on the age group pair.[151] For age group pairs 0 through 4 and 7 through 9, dichotomous logistic regression models were built. Because there were three outcomes with age group pairs 5, 6, and 10, polytomous logistic models were fitted for these age group pairs. All the models incorporated the weights and were calibrated to account for item nonresponse (where a pair responded to the survey but the pair relationship was unknown), using the item response propensity models, as described in Section 11.5.2.1.

Ideally, each type of pair relationship within an age group pair would constitute a response category in a multinomial response model. However, the number of cases corresponding to some pair relationships within each age group pair was very small. Hence, it was not feasible to fit multinomial models that cover all the possible pair relationships for a given age group pair. Therefore, in the modeling step, some of the response categories were combined with separate assignments of imputed values within each of the 11 age group pairs. Priority was given to placing the pair relationships "of interest" into separate categories. In some cases, pair relationships that were not of interest were combined with other categories, even if there were sufficient numbers to have a separate category in the multinomial model. Table 11.9 presents the response categories that were used for modeling. The delineation between categories that were combined for modeling was determined in the hot-deck step.

Table 11.9 Modeled Pair Relationships within Age Group Pairs

Age Group Pair Number	Age Group Pair	Number of Levels in Response	Levels of Modeled Response
0	12–14/12–14	2	Sibling-sibling; all others
1	12–14/15–17	2	Sibling-sibling; all others
2	12–14/18–25	2	Sibling-sibling; all others
3	15–17/15–17	2	Sibling-sibling; all others
4	15–17/18–25	2	Sibling-sibling; all others
5	18–20/18–25	3	Both spouse-spouse pair relationships;[1] all others
6	21–25/21–25	3	Both spouse-spouse pair relationships;[1] all others
7	12–14/26+	2	Parent-child; all others

[150] Partitioning the observations into two groups is similar to including an indicator variable in the PRD model. Though two models were fitted in practice, in theory this is similar to fitting only one PRD model. Thus, this is still considered the single RP/single PRD type of PMN.

[151] Though the SUDAAN procedure RLOGIST could have been used for all age group pairs except 5, 6, and 9, MULTILOG was used for all models for coding simplicity.

Table 11.9 Modeled Pair Relationships within Age Group Pairs (continued)

Age Group Pair Number	Age Group Pair	Number of Levels in Response	Levels of Modeled Response
8	15–17/26+	2	Parent-child; all others
9	18–20/26+	2	Parent-child; all others
10	21+/26+	3	Both spouse-spouse pair relationships;[1] all others

[1] The two spouse-spouse pair relationships are spouse-spouse and spouse-spouse-with-children-younger-than-18. The pair relationships labeled "spouse-spouse" include partner-partner pair relationships.

As an example, consider age group pair 5. In this age group pair, there are typically four types of pair relationships that have a sufficient number of respondent pairs to fit an adequate model, including both spouse-spouse domains, sibling-sibling pairs, and all others. Models with fewer response levels are generally easier to fit because there are more observations in each response level. Because only two of those four were pair relationships of interest, the two spouse-spouse domains were used as levels in the response variable. The third level was obtained by combining the sibling-sibling and other relationship pairs. There are typically a small number of parent-child pairs, which also were combined with the other relationship pairs.

11.5.2.3 Hot-Deck Step

Likeness constraints used in imputation of the pair relationship were generally based on the number of household members in various age groups and on the marital status and genders of the respondents. Logical constraints were limited to the information that was already known about the pair, as denoted by the level of the variable PAIRREL. If, for example, PAIRREL = 14, then no information was available about the identity of the pair relationship and no logical constraint was needed. On the other hand, if PAIRREL = 15, this meant that the pair relationship was either a parent-child pair where the child was aged 12 to 14 or it was some relationship other than spouse-spouse, parent-child, grandparent-grandchild, or sibling-sibling.

11.5.3 Stage Two Imputation: Multiplicity Counts

In many cases where the pair relationships were not defined, multiplicity counts also were not defined. In addition, there were a handful of cases where multiplicity counts were not determined, even when the pair relationship was known. In all of these cases, imputation was required to determine the multiplicity count, and the single RP/single PRD type of PMN was used for imputation.

The multiplicity count variables were divided into six imputation sets: four sets for the multiplicities associated with the four sibling-sibling pair domains; one set for the parent-child, child focus domains; and one set for the parent-child, parent focus domains. The variables in each imputation set are provided in Table 11.10. Because the parent-child counts are hierarchical (i.e., the count for 12-17 must be less than or equal to the count for 12-20), only the counts for the 12-20 age group were modeled. Using the predicted means from these models, a single donor pair for each focus was selected from which the multiplicity counts were determined for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs. No imputation was required for the spouse-

spouse multiplicity counts, because a selected respondent in a spouse-spouse pair was assumed to have had only one spouse.

Table 11.10 Modeled Multiplicities within Domains

Imputation Set	Domain	Model Type	Base Variables	Imputation-Revised Variables
1	Parent-Child, Child Focus	Logistic	MCPCC14, MCPCC57, MCPCC17, MCPCC20	IRMPCC14, IRMPCC57, IRMPCC17, IRMPCC20
2	Parent-Child, Parent Focus	Poisson	MCPCP14, MCPCP57, MCPCP17, MCPCP20	IRMPCP14, IRMPCP57, IRMPCP17, IRMPCP20
3	Sibling-Sibling (12-14/15-17) Older Focus	Poisson	MCS1417	IRMS1417
4	Sibling-Sibling (12-14/15-17) Younger Focus	Poisson	MCS1714	IRMS1714
5	Sibling-Sibling (12-17/18-25) Older Focus	Poisson	MCS1725	IRMS1725
6	Sibling-Sibling (12-17/18-25) Younger Focus	Poisson	MCS2517	IRMS2517

11.5.3.1 Imputation for Parent-Child Multiplicity, Child Focus (Imputation Set 1)

The first imputation set included the four parent-child, child focus multiplicity variables. Modeling was done only for the parent-child (12-20) variable, and multivariate assignment of all four variables was done in the hot-deck step.

11.5.3.1.1 Response Propensity Step

For a respondent pair to be considered complete with regard to the parent-child multiplicities, the multiplicity had to be nonmissing for the domains with children aged 12 to 20. A nonmissing multiplicity for this domain would automatically guarantee nonmissing multiplicities for the subset parent-child domains. Response propensity adjustments were then calculated in order to make the respondent pair weights representative of the entire sample of pairs.

11.5.3.1.2 Prediction Step

For the child-focus parent-child domains, the count being modeled was the number of parents of children aged 12 to 20 who were part of a parent-child pair. In most cases only two responses were possible: one parent or two parents. There were rare instances where the child could have three parents living in the household, with some combination of biological, step, foster, or adoptive parents. For the purposes of modeling, these cases were collapsed with the two-parent households. Similar to the procedure described in Section 11.5.2.2, the data were divided into two groups based on whether the age count variables were nonmissing, and separate models were fitted for each group. The fitted models were binomial logistic regression models using the SUDAAN procedure RLOGIST.

11.5.3.1.3 Hot-Deck Step

Though modeling was only done for the 12-20 age group, the multiplicity counts for the 12-14, 12-17, 15-17, and 12-20 parent-child pairs were assigned from a single donor pair. Likeness constraints were generally based on the pair relationship, the household size, and the number of household members in various age groups. Logical constraints were based on the bounds created during the editing process. These constraints are described in detail in Tables E.112 and E.113 in Appendix E.

11.5.3.2 Imputation for All Other Multiplicities (Imputation Sets 2 through 6)

The imputation process for the other responses (parent-focus parent-child and sibling-sibling multiplicity counts) was similar to that for the child-focus parent-child counts. The main difference is in the PRD step where Poisson regression models were used to model the counts. The counts of the number of children or siblings were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance.

As with the child-focus parent-child multiplicity counts, the parent-focus parent-child multiplicity counts were only modeled for the 12-20 age group, and counts from all four age groups were assigned from a single donor pair. For the sibling-sibling multiplicity counts, there was only one variable in each imputation set, and each count was modeled separately.

11.5.4 Stage Three Imputation: Household-Level Person Counts

Because of the difficulty in definitively determining household-level counts in many cases, imputation was not uncommon. Household-level person counts were divided into five imputation sets based on the domains listed in Section 11.4: one for each of the two sibling-sibling counts, one for the spouse-spouse counts, one for the spouse-spouse with children counts, and one for the parent-child counts. The first four imputation sets were handled using the single RP/single PRD type of PMN, whereas the imputation of the parent-child household counts was done using the multiple RP/multiple PRD type of PMN. For these counts, separate models were fit for the child-focus and parent-focus counts, and the predicted means from both models were brought together in one hot-deck step.

As with the multiplicities, the parent-child domains were hierarchical, so the imputations could not have been conducted independently if consistency was to be maintained. Hence, models were fitted only to the parent-child domains for the 12-20 age group, and the household-level person counts were assigned for the 12-14, 12-17, and 12-20 parent-child pair domains from a single donor. The household-level person counts for the 15-17 parent-child domains were not determined as they can be easily derived. The spouse-spouse household-level person counts were also hierarchical in that knowledge of whether a spouse-spouse pair was in the household was required before one could say that the pair had children. Therefore, imputations of the spouse-spouse counts were processed first, followed by the imputations of whether the spouse-spouse pairs in the household had children.

Household-level person counts were defined for all respondents, regardless of which pair they belonged to, or even whether they were within a pair at all. For modeling purposes,

respondents were partitioned into two groups based on whether they belonged to a pair, and the entire imputation process was conducted separately for each group.

11.5.4.1 Imputation for Sibling-Sibling (12-14/15-17), Older Focus Household Counts (Imputation Set 1)

The first imputation set included one base variable, HCS1417, and the corresponding imputation-revised variable IRHS1417. The imputation process for this variable is detailed in Sections 11.5.4.1.1 through 11.5.4.1.3.

11.5.4.1.1 Response Propensity Step

For a pair or single respondent to be considered complete, the household-level person counts had to be nonmissing for all the variables being imputed. In addition to being separated into pair and nonpair households, respondents also were split by age for both the modeling of the response propensity and the predicted means. For the pairs, households where both pair members were younger than 18 were placed in one group, and the remaining pairs were in the other age group. For the single respondents, one age group consisted of respondents who were younger than 18, and the other consisted of respondents who were 18 or older. For pairs, response propensity adjustments were calculated in order to make the household weights representative of the entire sample of pairs. For single respondents, household weights also were used, and the adjustments were calculated in order to make the respondent household weights representative of the entire sample of households that were not part of a pair.

11.5.4.1.2 Prediction Step

After the response propensity adjustment was complete, the data were split into two groups based on whether the household-level age count variables were nonmissing, as previously described in Section 11.5.2.2. The outcome variable for the household-level sibling-sibling (12-14/15-17) count models was the number of household members aged 15 to 17 with a sibling aged 12 to 14 living with them. These counts could have a value of zero, which distinguished them from the multiplicities from a modeling point of view. Poisson regression was used to fit the models for the household-level person counts corresponding to the sibling-sibling domains. The data were underdispersed for a Poisson distribution so that the data had to be scaled using the observed variance. Modeling was done using the SUDAAN procedure LOGLINK.

11.5.4.1.3 Hot-Deck Step

After the modeling steps were complete, the two age groups were combined for one hot-deck imputation step. Imputation was conducted separately for pair and nonpair households. Likeness constraints used in imputation of household counts were generally based on the number of families in the household, the household size, and the number of household members in various age groups. Logical constraints were based on the bounds created during the editing process. These constraints are described in detail in Tables E.121 and E.122 in Appendix E.

In those instances where an imputed value could not be found after loosening all the likeness constraints, the imputed value was determined by doing a random imputation within bounds derived from the household composition.

11.5.4.2 Imputation for Sibling-Sibling (12-17/18-25) and Both Spouse-Spouse Household Counts (Imputation Sets 2 through 4)

The second, third, and fourth imputation sets were processed similarly to the first. One large difference was that when modeling the spouse-spouse-with-children counts, the data were not separated according to age groups. This applied to both the response propensity adjustments and the calculation of predicted means. Because of the hierarchical relationship between the spouse-spouse and the spouse-spouse-with-children counts, the response propensity adjustment for the spouse-spouse-with-children domain adjusted the weights to be representative of all spouse-spouse pairs rather than the entire sample. Missing counts for the spouse-spouse-with-children domain were not replaced via imputation until it was known definitively, after the hot-deck step of the PMN imputation, whether a household had spouse-spouse pairs.

Polytomous logistic regression was used to model the count of spouse-spouse pairs, with the possible responses being zero, one, and two or more spouse-spouse pairs in the household. Whether or not the spouse-spouse pairs had children younger than 18 was modeled with binomial logistic regression. These models were fitted using the SUDAAN MULTILOG procedure.

In some cases, two family units were in a household. If these resulted in unusual household-level person counts, they were excluded from the modeling step and were considered nonrespondents for the purposes of weight adjustment. No predicted mean was calculated in these cases, and instead of matching donors and recipients using predicted means, the imputed value was determined using random imputation within the preset bounds. One case where this may have occurred was with the spouse-spouse-with-children counts. Having two spouse-spouse pairs with children younger than 18 was an extremely rare category. Therefore, the two response categories that resulted for the spouse-spouse-with-children models were zero or one or more. Households with two family units did not need to be excluded from the spouse-spouse models, because having two spouse-spouse pairs in a household, though not common, was not rare.

11.5.4.3 Imputation for Parent-Child Household Counts (Imputation Set 5)

In contrast to the first four imputation sets, the parent-child household counts were imputed using the multiple RP/multiple PRD type of PMN. This imputation set included the three child-focus counts, HCPCC14, HCPCC17, and HCPCC20, as well as the three parent-focus counts, HCPCP14, HCPCP17, and HCPCP20. The corresponding imputation-revised variables are IRHPCC14, IRHPCC17, IRHPCC20, IRHPCP14, IRHPCP17, and IRHPCP20, respectively. The child-focus and parent-focus counts were modeled separately and joined together in one final hot-deck step. Just as was done for the first four imputation sets, the respondents were grouped into pair and nonpair households, and the entire imputation process was completed separately for each group. Unlike most imputation sets processed using the multiple RP/multiple PRD type of PMN, no provisional hot-deck steps were implemented. Because the child-focus counts were not used as covariates for the parent-focus counts and the parent-focus counts were not used as covariates for the child-focus counts, no provisional hot-deck steps were necessary, and the RP and PRD steps for the child-focus counts were run in parallel with the RP and PRD steps for the parent-focus counts.

11.5.4.3.1 First Response Propensity Step: Child Focus

All respondents were in the domain for all household counts. For a pair or single respondent to be considered complete, the three child-focus counts had to be nonmissing. Similar to the processing of the sibling-sibling counts and the spouse-spouse counts, the data were divided into two age groups based on whether the respondent or pair of respondents was 18 or older.

11.5.4.3.2 First Prediction Step: Child Focus

For the child-focus counts, the modeled response was the number of children aged 12 to 20 in the household with at least one parent living with them. This was modeled using Poisson regression, where the data were scaled using the observed variance to account for underdispersion. The LOGLINK procedure in SUDAAN was used to fit the model. The predicted mean used in the final hot-deck step was the predicted number of children aged 12 to 20 in the household with at least one parent living with them.

11.5.4.3.3 Second Response Propensity Step: Parent Focus

This RP step was identical to the RP step for the child-focus counts, except item respondents were those whose three parent-focus counts were nonmissing.

11.5.4.3.4 Second Prediction Step: Parent Focus

For the parent-focus counts, the modeled response was a three-level variable, based on the number of parents in the household with children aged 12 to 20: zero, one, or two or more. Polytomous logistic regression was used to fit the model as implemented by the MULTILOG procedure in SUDAAN. The predicted means used in the final hot-deck step were the predicted probabilities associated with each of the three levels of the response variable.

11.5.4.3.5 Hot-Deck Step

The predicted means from the models fit for the child-focus and parent-focus counts were brought together in one hot-deck imputation step. Because there are six variables in the imputation set, and any combination of them could be missing, there were 63 possible missingness patterns. These patterns are enumerated in Table E.127 in Appendix E. As with the other four imputation sets, logical constraints were based on the bounds created during the editing process, and likeness constraints were generally based on the number of families in the household, the household size, and the number of household members in various age groups. These constraints are described in detail in Tables E.121 and E.122.

In cases where there were two family units in the household, resulting in unusual counts, the counts were not included in the parent-focus models, and no predicted means were calculated. Even though two-family households were included in the model for the child-focus parent-child counts, the resulting predicted means were not used. This was because the parent-focus and child-focus parent-child counts were in the same imputation set, and the predicted means could not be used in the imputation of the parent-focus parent-child counts when two families were in the household. In these cases, imputation was random between the bounds.

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th ed.). Washington, DC: Author.

Ault, K., Aldworth, J., Barnett-Walker, K., Carpenter, L., Copello, E., Frechtel, P., Liu, B., & Martin, P. (2009). Imputation report. In *2007 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.377.007). Research Triangle Park, NC: RTI International.

Ault, K., Barnett-Walker, K., Carpenter, L., Copello, E., Cummiskey, C., Frechtel, P., Laufenberg, J., Liu, B., Martin, P., & Moore, A. (2010). Imputation report. In *2008 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.477.007). Research Triangle Park, NC: RTI International.

Ault, K., Barnett-Walker, K., Carpenter, L., Cummiskey, C., Frechtel, P., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2011). Imputation report. In *2009 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009.577.007). Research Triangle Park, NC: RTI International.

Ault, K., Barnett-Walker, K., Carpenter, L., Cummiskey, C., Dai, L., Edwards, S., Eicheldinger, C., Frechtel, P., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (in press). *National Survey on Drug Use and Health: Evaluation of imputation methods* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, RTI/0211838.108.006.018). Research Triangle Park, NC: RTI International.

Center for Behavioral Health Statistics and Quality. (2013a). *Results from the 2012 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2013b). *Results from the 2012 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 13-4795, NSDUH Series H-46). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 41, RTI/0212800.001.107.004). Research Triangle Park, NC: RTI International.

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Cox, B. G. (1980). The weighted sequential hot deck imputation procedure. In *Proceedings of the 1980 American Statistical Association, Survey Research Methods Section, Houston, TX* (pp. 721-726). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Cox, B. G., & Cohen, S. B. (1985). *Methodological issues for health care surveys*. New York: Marcel Dekker, Inc.

Cox, D. R., & Snell, E. J. (1989). *The analysis of binary data* (2nd ed.). (Monographs on Statistics and Applied Probability Series). Boca Raton, FL: CRC Press.

Frechtel, P., Archambault, H., Carpenter, L., Cummiskey, C., Edwards, S., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2013). Imputation report. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.006.007). Research Triangle Park, NC: RTI International.

Iannacchione, V. (1982). Weighted sequential hot deck imputation macros. In *Proceedings of the Seventh Annual SAS Users Group International Conference* (pp. 759-763). Cary, NC: SAS Corporation.

Kroutil, L. A., Chien, C., Handley, W., & Bradshaw, M. R. (2013). Procedures for editing interviewer-administered data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International.

Kroutil, L. A., Handley, W., & Bradshaw, M. R. (2013). General principles and procedures for editing drug use data in the 2011 NSDUH computer-assisted interview. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 10, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.207.003). Research Triangle Park, NC: RTI International.

Kroutil, L., & Myers, L. (2002). Development of editing rules for CAI substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (HHS Publication No. SMA 03-3768, pp. 85-109). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

Office of Applied Studies. (2008). *Results from the 2007 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 08-4343, NSDUH Series H-34). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2009). *Results from the 2008 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010a). *Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of national findings* (HHS Publication No. SMA 10-4586Findings, NSDUH Series H-38A). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010b). *Results from the 2009 National Survey on Drug Use and Health: Volume II. Technical appendices and selected prevalence tables* (HHS Publication No. SMA 10-4586Appendices, NSDUH Series H-38B). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register, 62*(210), 58781-58790.

RTI International. (2008). *SUDAAN® language manual, Release 10.0*. Research Triangle Park, NC: RTI International.

Rubin, D. B. (1986). Statistical matching using file concatenation with adjusted weights and multiple imputations. *Journal of Business and Economic Statistics, 4*(1), 87-94.

Ruppenkamp, J., Davis, T., Kroutil, L., & Aldworth, J. (2006). Methamphetamine analysis report. In *2005 National Survey on Drug Use and Health: Methodological resource book* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, RTI/0208726.187.038). Research Triangle Park, NC: RTI International.

Ruppenkamp, J., Frechtel, P., Aldworth, W. J., Carpenter, L., Clarke, A., Davis, T., Kroutil, L. A, & Martin, P. C. (2007). Methamphetamine analysis report. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 15, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Deliverable No. 39, RTI/0209009). Research Triangle Park, NC: RTI International.

SAS Institute Inc. (2013). *SAS/STAT® 13.1 user's guide.* Cary, NC: SAS Institute Inc.

Schafer, J. L. (1997). *Analysis of incomplete multivariate data* (No. 72, Monographs on Statistics and Applied Probability). Boca Raton, FL: Chapman and Hall/CRC.

Scott, V., Archambault, H., Carpenter, L., Edwards, S., Frechtel, P., King, S., Laufenberg, J., & Moore, A. (2013). *2012 National Survey on Drug Use and Health: Editing and imputation evaluation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 15, RTI/0212800.001.108.003). Research Triangle Park, NC: RTI International.

Shiffman, S., Hickcox, M., Gnys, M., Paty, J. A., & Kassel, J. D. (1995, March). *The Nicotine Dependence Syndrome Scale: Development of a new measure.* Poster presented at the annual meeting of the Society for Research on Nicotine and Tobacco, San Diego, CA.

Shiffman, S., Waters, A. J., & Hickcox, M. (2003). The Nicotine Dependence Syndrome Scale: A multi-dimensional measure of nicotine dependence. Unpublished manuscript.

Singh, A., Grau, E., & Folsom, R. (2004). Imputation and unbiased estimation: Use of centered predictive mean neighborhoods method. In *Proceedings of the 2004 Joint Statistical Meetings, American Statistical Association, Section on Survey Research Methods, Toronto, Ontario, Canada* (pp. 4351-4358). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Westlake, M., Chen, P., & Gordek, H. (2014). Questionnaire dwelling unit-level and person pair-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 13, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.107.005). Research Triangle Park, NC: RTI International.

Appendix A: Percentage of Cases Imputed or Logically Assigned, 2012 Survey

Table A.1 Percentage of Cases Imputed for General Cigarette Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Cigarette Recency	Lifetime Cigarette Users	33,970	426	5	431	0.35%	0.01%	0.36%
Cigarette Frequency Past Month	Past Month Cigarette Users	15,062	31	123	154	0.11%	1.04%	1.15%
Cigarette Age at First Use	Lifetime Cigarette Users	33,970	444	13	457	0.95%	0.04%	0.99%
Cigarette Day of First Use	Lifetime Cigarette Users	33,970	33,970	0	33,970	100.00%	0.00%	100.00%
Cigarette Month of First Use	Lifetime Cigarette Users	33,970	31,585	21	31,606	97.98%	0.03%	98.01%
Cigarette Year of First Use	Lifetime Cigarette Users	33,970	31,433	5	31,438	97.87%	0.00%	97.87%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.2 Percentage of Cases Imputed for Daily Cigarette Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Cigarette Daily Use	Lifetime Cigarette Users	33,970	22	43	65	0.07%	0.18%	0.24%
Cigarette Age at First Daily Use	Daily Cigarette Users	16,929	118	1	119	0.98%	0.00%	0.98%
Cigarette Day of First Daily Use	Daily Cigarette Users	16,929	16,929	0	16,929	100.00%	0.00%	100.00%
Cigarette Month of First Daily Use	Daily Cigarette Users	16,929	16,079	0	16,079	98.69%	0.00%	98.69%
Cigarette Year of First Daily Use	Daily Cigarette Users	16,929	16,027	1	16,028	98.61%	0.00%	98.61%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.3 Percentage of Cases Imputed for Nicotine Dependency Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Smoke Cigarettes Regularly throughout the Day	Past Month Cigarette Users	15,062	0	0	0	0.00%	0.00%	0.00%
Smoke Same Number of Cigarettes from Day to Day	Past Month Cigarette Users	15,062	5	0	5	0.06%	0.00%	0.06%
Number Cigarettes Smoked per Day Often Changes	Past Month Cigarette Users	15,062	10	0	10	0.05%	0.00%	0.05%
Number Cigarettes Smoked per Day Influenced by Other Things	Past Month Cigarette Users	15,062	4	0	4	0.02%	0.00%	0.02%
Smoke Same Number of Cigarettes on Weekends As on Weekdays	Past Month Cigarette Users	15,062	9	0	9	0.04%	0.00%	0.04%
Smoke Lots of Cigarettes in an Hour, Then No Cigarettes Until Hours Later	Past Month Cigarette Users	15,062	20	0	20	0.07%	0.00%	0.07%
Amount of Smoking Has Increased Since Started Smoking	Past Month Cigarette Users	15,062	1	0	1	0.00%	0.00%	0.00%
Must Smoke Much More Now Before Start To Feel Anything	Past Month Cigarette Users	15,062	21	0	21	0.09%	0.00%	0.09%
Need To Smoke a Lot More To Be Satisfied	Past Month Cigarette Users	15,062	3	0	3	0.01%	0.00%	0.01%
Crave Cigarettes when Haven't Smoked for a Few Hours	Past Month Cigarette Users	15,062	2	0	2	0.03%	0.00%	0.03%
Need To Smoke To Feel Less Irritable	Past Month Cigarette Users	15,062	9	0	9	0.18%	0.00%	0.18%
Feel in Control of Smoking	Past Month Cigarette Users	15,062	7	0	7	0.03%	0.00%	0.03%
Cravings for Cigarettes like Force Can't Control	Past Month Cigarette Users	15,062	6	0	6	0.04%	0.00%	0.04%
Worry about Running Out of Cigarettes	Past Month Cigarette Users	15,062	0	0	0	0.00%	0.00%	0.00%

Table A.3 Percentage of Cases Imputed for Nicotine Dependency Variables (continued)

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Smoking Not Affected by Other Things	Past Month Cigarette Users	15,062	11	0	11	0.03%	0.00%	0.03%
Tend To Avoid Places That Don't Allow Smoking	Past Month Cigarette Users	15,062	23	0	23	0.09%	0.00%	0.09%
No Travel by Airplane Because No Smoking Allowed	Past Month Cigarette Users	15,062	20	0	20	0.11%	0.00%	0.11%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.4 Percentage of Cases Imputed for Cigar Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Cigar Lifetime Use	All Respondents	68,309	16	0	16	0.03%	0.00%	0.03%
Cigar Recency	Lifetime Cigar Users	20,204	364	3	367	0.69%	0.01%	0.70%
Cigar Frequency Past Month	Past Month Cigar Users	4,252	16	35	51	0.37%	0.56%	0.93%
Cigar Age at First Use	Lifetime Cigar Users	20,204	439	0	439	2.22%	0.00%	2.22%
Cigar Day of First Use	Lifetime Cigar Users	20,204	20,204	0	20,204	100.00%	0.00%	100.00%
Cigar Month of First Use	Lifetime Cigar Users	20,204	17,798	11	17,809	95.85%	0.02%	95.87%
Cigar Year of First Use	Lifetime Cigar Users	20,204	17,638	0	17,638	95.59%	0.00%	95.59%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.5 Percentage of Cases Imputed for Chewing Tobacco Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Chewing Tobacco Lifetime Use	All Respondents	68,309	8	0	8	0.01%	0.00%	0.01%
Chewing Tobacco Recency	Lifetime Chewing Tobacco Users	7,383	148	0	148	0.83%	0.00%	0.83%
Chewing Tobacco Frequency Past Month	Past Month Chewing Tobacco Users	1,084	4	16	20	0.25%	0.72%	0.97%
Chewing Tobacco Age at First Use	Lifetime Chewing Tobacco Users	7,383	158	0	158	1.51%	0.00%	1.51%
Chewing Tobacco Day of First Use	Lifetime Chewing Tobacco Users	7,383	7,383	0	7,383	100.00%	0.00%	100.00%
Chewing Tobacco Month of First Use	Lifetime Chewing Tobacco Users	7,383	6,788	4	6,792	97.66%	0.01%	97.67%
Chewing Tobacco Year of First Use	Lifetime Chewing Tobacco Users	7,383	6,730	0	6,730	97.43%	0.00%	97.43%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.6 Percentage of Cases Imputed for Snuff Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Snuff Lifetime Use	All Respondents	68,309	27	0	27	0.06%	0.00%	0.06%
Snuff Recency	Lifetime Snuff Users	9,377	220	0	220	1.08%	0.00%	1.08%
Snuff Frequency Past Month	Past Month Snuff Users	2,467	6	21	27	0.12%	0.54%	0.66%
Snuff Age at First Use	Lifetime Snuff Users	9,377	206	0	206	1.74%	0.00%	1.74%
Snuff Day of First Use	Lifetime Snuff Users	9,377	9,377	0	9,377	100.00%	0.00%	100.00%
Snuff Month of First Use	Lifetime Snuff Users	9,377	8,380	6	8,386	96.19%	0.02%	96.21%
Snuff Year of First Use	Lifetime Snuff Users	9,377	8,303	0	8,303	95.99%	0.00%	95.99%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.7 Percentage of Cases Imputed for Smokeless Tobacco Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Smokeless Tobacco Lifetime Use	All Respondents	68,309	27	0	27	0.06%	0.00%	0.06%
Smokeless Tobacco Recency	Lifetime Smokeless Tobacco Users	11,353	295	0	295	1.18%	0.00%	1.18%
Smokeless Tobacco Age at First Use	Lifetime Smokeless Tobacco Users	11,353	317	0	317	2.03%	0.00%	2.03%
Smokeless Tobacco Day of First Use	Lifetime Smokeless Tobacco Users	11,353	11,353	0	11,353	100.00%	0.00%	100.00%
Smokeless Tobacco Month of First Use	Lifetime Smokeless Tobacco Users	11,353	10,257	5	10,262	96.85%	0.01%	96.86%
Smokeless Tobacco Year of First Use	Lifetime Smokeless Tobacco Users	11,353	10,168	0	10,168	96.64%	0.00%	96.64%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.8 Percentage of Cases Imputed for Pipe Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Pipe Lifetime Use	All Respondents	68,309	9	0	9	0.01%	0.00%	0.01%
Pipe Past Month Use	Lifetime Pipe Users	5,888	3	0	3	0.08%	0.00%	0.08%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.9 Percentage of Cases Imputed for Alcohol Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Alcohol Lifetime Use	All Respondents	68,309	23	0	23	0.03%	0.00%	0.03%
Alcohol Recency	Lifetime Alcohol Users	47,495	853	12	865	1.21%	0.01%	1.21%
Alcohol Frequency Past Year	Past Year Alcohol Users	40,595	800	1,435	2,235	1.41%	2.05%	3.46%
Alcohol Frequency Past Month	Past Month Alcohol Users	30,218	646	192	838	1.74%	0.53%	2.28%
Alcohol 5+ Drinks Past Month	Past Month Alcohol Users	30,218	1,035	80	1,115	3.19%	0.27%	3.46%
Alcohol Age at First Use	Lifetime Alcohol Users	47,495	645	0	645	1.31%	0.00%	1.31%
Alcohol Day of First Use	Lifetime Alcohol Users	47,495	47,495	0	47,495	100.00%	0.00%	100.00%
Alcohol Month of First Use	Lifetime Alcohol Users	47,495	42,530	9	42,539	96.95%	0.00%	96.95%
Alcohol Year of First Use	Lifetime Alcohol Users	47,495	42,227	0	42,227	96.78%	0.00%	96.78%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.10 Percentage of Cases Imputed for Marijuana Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Marijuana Lifetime Use	All Respondents	68,309	45	0	45	0.06%	0.00%	0.06%
Marijuana Recency	Lifetime Marijuana Users	27,199	452	4	456	0.93%	0.00%	0.93%
Marijuana Frequency Past Year	Past Year Marijuana Users	12,657	428	945	1,373	2.70%	4.90%	7.61%
Marijuana Frequency Past Month	Past Month Marijuana Users	7,425	197	69	266	2.34%	0.63%	2.97%
Marijuana Age at First Use	Lifetime Marijuana Users	27,199	210	0	210	0.71%	0.00%	0.71%
Marijuana Day of First Use	Lifetime Marijuana Users	27,199	27,199	0	27,199	100.00%	0.00%	100.00%
Marijuana Month of First Use	Lifetime Marijuana Users	27,199	24,570	11	24,581	96.77%	0.02%	96.79%
Marijuana Year of First Use	Lifetime Marijuana Users	27,199	24,417	0	24,417	96.59%	0.00%	96.59%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.11 Percentage of Cases Imputed for Inhalant Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Inhalant Lifetime Use	All Respondents	68,309	124	0	124	0.08%	0.00%	0.08%
Inhalant Recency	Lifetime Inhalant Users	5,654	227	4	231	1.94%	0.04%	1.98%
Inhalant Frequency Past Year	Past Year Inhalant Users	975	102	120	222	8.19%	7.88%	16.07%
Inhalant Frequency Past Month	Past Month Inhalant Users	257	31	10	41	8.62%	2.35%	10.98%
Inhalant Age at First Use	Lifetime Inhalant Users	5,654	293	0	293	3.21%	0.00%	3.21%
Inhalant Day of First Use	Lifetime Inhalant Users	5,654	5,654	0	5,654	100.00%	0.00%	100.00%
Inhalant Month of First Use	Lifetime Inhalant Users	5,654	5,078	3	5,081	96.55%	0.03%	96.58%
Inhalant Year of First Use	Lifetime Inhalant Users	5,654	4,996	0	4,996	96.14%	0.00%	96.14%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.12 Percentage of Cases Imputed for Heroin Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Heroin Lifetime Use	All Respondents	68,309	51	0	51	0.05%	0.00%	0.05%
Heroin Recency	Lifetime Heroin Users	952	31	1	32	1.80%	0.08%	1.88%
Heroin Frequency Past Year	Past Year Heroin Users	249	18	38	56	6.91%	11.64%	18.55%
Heroin Frequency Past Month	Past Month Heroin Users	102	9	0	9	7.70%	0.00%	7.70%
Heroin Age at First Use	Lifetime Heroin Users	952	12	0	12	1.47%	0.00%	1.47%
Heroin Day of First Use	Lifetime Heroin Users	952	952	0	952	100.00%	0.00%	100.00%
Heroin Month of First Use	Lifetime Heroin Users	952	837	0	837	94.61%	0.00%	94.61%
Heroin Year of First Use	Lifetime Heroin Users	952	834	0	834	94.58%	0.00%	94.58%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.13 Percentage of Cases Imputed for Hallucinogen Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Hallucinogen Lifetime Use	All Respondents	68,309	233	0	233	0.20%	0.00%	0.20%
Hallucinogen Recency	Lifetime Hallucinogen Users	8,780	241	50	291	1.34%	0.31%	1.65%
Hallucinogen Frequency Past Year	Past Year Hallucinogen Users	2,178	107	142	249	3.25%	4.81%	8.05%
Hallucinogen Frequency Past Month	Past Month Hallucinogen Users	578	46	3	49	7.16%	0.28%	7.44%
Hallucinogen Age at First Use	Lifetime Hallucinogen Users	8,780	122	98	220	1.29%	1.03%	2.32%
Hallucinogen Day of First Use	Lifetime Hallucinogen Users	8,780	8,780	0	8,780	100.00%	0.00%	100.00%
Hallucinogen Month of First Use	Lifetime Hallucinogen Users	8,780	7,726	62	7,788	95.86%	0.21%	96.07%
Hallucinogen Year of First Use	Lifetime Hallucinogen Users	8,780	7,675	19	7,694	95.66%	0.06%	95.72%
Hallucinogen Use Other than LSD, PCP, or Ecstasy	All Respondents	68,309	207	0	207	0.19%	0.00%	0.19%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.14 Percentage of Cases Imputed for LSD Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
LSD Lifetime Use	All Respondents	68,309	61	0	61	0.06%	0.00%	0.06%
LSD Recency	Lifetime LSD Users	4,382	84	0	84	0.66%	0.00%	0.66%
LSD Age at First Use	Lifetime LSD Users	4,382	87	5	92	1.27%	0.05%	1.32%
LSD Day of First Use	Lifetime LSD Users	4,382	4,382	0	4,382	100.00%	0.00%	100.00%
LSD Month of First Use	Lifetime LSD Users	4,382	3,999	0	3,999	97.65%	0.00%	97.65%
LSD Year of First Use	Lifetime LSD Users	4,382	3,980	0	3,980	97.59%	0.00%	97.59%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.15 Percentage of Cases Imputed for PCP Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
PCP Lifetime Use	All Respondents	68,309	63	0	63	0.05%	0.00%	0.05%
PCP Recency	Lifetime PCP Users	988	35	0	35	1.78%	0.00%	1.78%
PCP Age at First Use	Lifetime PCP Users	988	35	5	40	2.57%	0.22%	2.78%
PCP Day of First Use	Lifetime PCP Users	988	988	0	988	100.00%	0.00%	100.00%
PCP Month of First Use	Lifetime PCP Users	988	908	0	908	97.84%	0.00%	97.84%
PCP Year of First Use	Lifetime PCP Users	988	903	0	903	97.75%	0.00%	97.75%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.16 Percentage of Cases Imputed for Ecstasy Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Ecstasy Lifetime Use	All Respondents	68,309	84	0	84	0.08%	0.00%	0.08%
Ecstasy Recency	Lifetime Ecstasy Users	4,817	116	0	116	1.46%	0.00%	1.46%
Ecstasy Age at First Use	Lifetime Ecstasy Users	4,817	60	0	60	0.99%	0.00%	0.99%
Ecstasy Day of First Use	Lifetime Ecstasy Users	4,817	4,817	0	4,817	100.00%	0.00%	100.00%
Ecstasy Month of First Use	Lifetime Ecstasy Users	4,817	4,014	0	4,014	91.87%	0.00%	91.87%
Ecstasy Year of First Use	Lifetime Ecstasy Users	4,817	3,977	0	3,977	91.49%	0.00%	91.49%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.17 Percentage of Cases Imputed for Cocaine Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Cocaine Lifetime Use	All Respondents	68,309	20	0	20	0.03%	0.00%	0.03%
Cocaine Recency	Lifetime Cocaine Users	7,190	89	33	122	0.54%	0.40%	0.94%
Cocaine Frequency Past Year	Past Year Cocaine Users	1,512	67	137	204	5.37%	7.02%	12.39%
Cocaine Frequency Past Month	Past Month Cocaine Users	432	33	15	48	10.71%	6.32%	17.03%
Cocaine Age at First Use	Lifetime Cocaine Users	7,190	67	96	163	0.99%	1.15%	2.15%
Cocaine Day of First Use	Lifetime Cocaine Users	7,190	7,190	0	7,190	100.00%	0.00%	100.00%
Cocaine Month of First Use	Lifetime Cocaine Users	7,190	6,608	2	6,610	97.42%	0.00%	97.42%
Cocaine Year of First Use	Lifetime Cocaine Users	7,190	6,575	0	6,575	97.31%	0.00%	97.31%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.18 Percentage of Cases Imputed for Crack Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Crack Lifetime Use	All Respondents	68,309	25	0	25	0.04%	0.00%	0.04%
Crack Recency	Lifetime Crack Users	1,662	18	0	18	0.79%	0.00%	0.79%
Crack Frequency Past Year	Past Year Crack Users	208	10	16	26	6.15%	2.85%	9.00%
Crack Frequency Past Month	Past Month Crack Users	69	4	1	5	6.69%	2.19%	8.88%
Crack Age at First Use	Lifetime Crack Users	1,662	18	0	18	1.14%	0.00%	1.14%
Crack Day of First Use	Lifetime Crack Users	1,662	1,662	0	1,662	100.00%	0.00%	100.00%
Crack Month of First Use	Lifetime Crack Users	1,662	1,590	0	1,590	98.68%	0.00%	98.68%
Crack Year of First Use	Lifetime Crack Users	1,662	1,587	0	1,587	98.67%	0.00%	98.67%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.19 Percentage of Cases Imputed for Tranquilizer Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Tranquilizer Lifetime Use	All Respondents	68,309	137	0	137	0.10%	0.00%	0.10%
Tranquilizer Recency	Lifetime Tranquilizer Users	5,793	128	1	129	1.58%	0.00%	1.58%
Tranquilizer Frequency Past Year	Past Year Tranquilizer Users	1,933	89	143	232	3.50%	4.66%	8.16%
Tranquilizer Age at First Use	Lifetime Tranquilizer Users	5,793	173	0	173	3.27%	0.00%	3.27%
Tranquilizer Day of First Use	Lifetime Tranquilizer Users	5,793	5,793	0	5,793	100.00%	0.00%	100.00%
Tranquilizer Month of First Use	Lifetime Tranquilizer Users	5,793	4,986	4	4,990	92.60%	0.02%	92.62%
Tranquilizer Year of First Use	Lifetime Tranquilizer Users	5,793	4,935	0	4,935	92.27%	0.00%	92.27%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.20 Percentage of Cases Imputed for Sedative Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Sedative Lifetime Use	All Respondents	68,309	166	0	166	0.11%	0.00%	0.11%
Sedative Recency	Lifetime Sedative Users	1,132	33	1	34	0.71%	0.02%	0.73%
Sedative Frequency Past Year	Past Year Sedative Users	216	19	24	43	2.68%	6.33%	9.01%
Sedative Age at First Use	Lifetime Sedative Users	1,132	66	0	66	3.21%	0.00%	3.21%
Sedative Day of First Use	Lifetime Sedative Users	1,132	1,132	0	1,132	100.00%	0.00%	100.00%
Sedative Month of First Use	Lifetime Sedative Users	1,132	1,021	1	1,022	97.30%	0.01%	97.31%
Sedative Year of First Use	Lifetime Sedative Users	1,132	1,014	0	1,014	97.18%	0.00%	97.18%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.21 Percentage of Cases Imputed for Pain Reliever Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Pain Reliever Lifetime Use	All Respondents	68,309	250	0	250	0.17%	0.00%	0.17%
Pain Reliever Recency	Lifetime Pain Reliever Users	10,518	414	12	426	2.91%	0.03%	2.94%
Pain Reliever Frequency Past Year	Past Year Pain Reliever Users	4,426	329	328	657	6.65%	4.32%	10.97%
Pain Reliever Age at First Use	Lifetime Pain Reliever Users	10,518	596	16	612	5.70%	0.14%	5.83%
Pain Reliever Day of First Use	Lifetime Pain Reliever Users	10,518	10,518	0	10,518	100.00%	0.00%	100.00%
Pain Reliever Month of First Use	Lifetime Pain Reliever Users	10,518	9,114	18	9,132	93.68%	0.05%	93.72%
Pain Reliever Year of First Use	Lifetime Pain Reliever Users	10,518	8,997	1	8,998	93.10%	0.01%	93.11%
Pain Reliever Use Other than OxyContin	All Respondents	68,309	279	0	279	0.18%	0.00%	0.18%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.22 Percentage of Cases Imputed for OxyContin Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
OxyContin Lifetime Use	All Respondents	68,309	180	0	180	0.14%	0.00%	0.14%
OxyContin Recency	Lifetime OxyContin Users	2,318	103	3	106	3.22%	0.03%	3.25%
OxyContin Frequency Past Year	Past Year OxyContin Users	650	51	51	102	5.62%	5.26%	10.88%
OxyContin Age at First Use	Lifetime OxyContin Users	2,318	78	0	78	3.10%	0.00%	3.10%
OxyContin Day of First Use	Lifetime OxyContin Users	2,318	2,318	0	2,318	100.00%	0.00%	100.00%
OxyContin Month of First Use	Lifetime OxyContin Users	2,318	2,001	0	2,001	92.06%	0.00%	92.06%
OxyContin Year of First Use	Lifetime OxyContin Users	2,318	1,983	0	1,983	91.71%	0.00%	91.71%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.23 Percentage of Cases Imputed for Stimulant Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Core Plus Noncore Stimulant Lifetime Use	All Respondents	68,309	137	0	137	0.11%	0.00%	0.11%
Core Plus Noncore Stimulant Recency	Lifetime Core Plus Noncore Stimulant Users	4,882	109	11	120	1.74%	0.14%	1.88%
Stimulant Lifetime Use	All Respondents	68,309	138	0	138	0.11%	0.00%	0.11%
Stimulant Recency	Lifetime Stimulant Users	4,466	91	11	102	1.04%	0.16%	1.20%
Stimulant Frequency Past Year	Past Year Stimulant Users	1,283	52	134	186	3.99%	8.91%	12.90%
Stimulant Age at First Use	Lifetime Stimulant Users	4,466	109	22	131	1.77%	0.67%	2.43%
Stimulant Day of First Use	Lifetime Stimulant Users	4,466	4,466	0	4,466	100.00%	0.00%	100.00%
Stimulant Month of First Use	Lifetime Stimulant Users	4,466	3,928	10	3,938	95.23%	0.06%	95.29%
Stimulant Year of First Use	Lifetime Stimulant Users	4,466	3,892	1	3,893	94.95%	0.00%	94.96%
Stimulant Use Other than Methamphetamine	All Respondents	68,309	134	0	134	0.11%	0.00%	0.11%

Note: The estimates for stimulant lifetime use and stimulant recency include data from the methamphetamine items added in 2005 and 2006, but other estimates in this table do not.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.24 Percentage of Cases Imputed for Methamphetamine Use Variables

		Unweighted Frequencies				Weighted Percentages		
	Domain	**Respondents in Domain**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**	**Imputed**	**Logically Assigned**	**Imputed or Logically Assigned**
Core Plus Noncore Methamphetamine Lifetime Use	All Respondents	68,309	59	0	59	0.04%	0.00%	0.04%
Core Plus Noncore Methamphetamine Recency	Lifetime Core Plus Noncore Methamphetamine Users	2,311	58	0	58	2.13%	0.00%	2.13%
Methamphetamine Lifetime Use	All Respondents	68,309	59	0	59	0.04%	0.00%	0.04%
Methamphetamine Recency	Lifetime Methamphetamine Users	1,724	37	0	37	0.97%	0.00%	0.97%
Methamphetamine Frequency Past Year	Past Year Methamphetamine Users	269	19	31	50	8.72%	6.24%	14.96%
Methamphetamine Age at First Use	Lifetime Methamphetamine Users	1,724	57	0	57	2.45%	0.00%	2.45%
Methamphetamine Day of First Use	Lifetime Methamphetamine Users	1,724	1,724	0	1,724	100.00%	0.00%	100.00%
Methamphetamine Month of First Use	Lifetime Methamphetamine Users	1,724	1,607	0	1,607	97.87%	0.00%	97.87%
Methamphetamine Year of First Use	Lifetime Methamphetamine Users	1,724	1,601	0	1,601	97.81%	0.00%	97.81%

Note: The estimates for methamphetamine lifetime use and methamphetamine recency include data from the methamphetamine items added in 2005 and 2006, but other estimates in this table do not.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.25 Percentage of Cases Imputed for Demographic Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
General								
Education Level	All Respondents	68,309	9	0	9	0.01%	0.00%	0.01%
Marital Status	Respondents at Least 15 Years of Age	57,151	8	0	8	0.01%	0.00%	0.01%
Employment Status	Respondents at Least 15 Years of Age	57,151	45	0	45	0.07%	0.00%	0.07%
Employment Status: 18+	Respondents at Least 18 Years of Age	45,836	40	0	40	0.08%	0.00%	0.08%
Immigrant Status								
Born in the United States	All Respondents	68,309	16	6	22	0.01%	0.01%	0.02%
Immigrant Age of Entry in the United States	Respondents Not Born in the United States	7,111	15	0	15	0.14%	0.00%	0.14%
Hispanic or Latino Origin								
Hispanic or Latino Origin	All Respondents	68,309	162	4	166	0.11%	0.01%	0.11%
Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,483	68	81	149	0.36%	0.83%	1.19%
Single/Multiple Hispanic or Latino Origin Group	Hispanic or Latino Respondents	11,483	75	85	160	0.53%	0.84%	1.36%
Race								
Native Hawaiian	All Respondents	68,309	2,388	0	2,388	3.48%	0.00%	3.48%
Other Pacific Islander	All Respondents	68,309	2,388	1	2,389	3.48%	0.00%	3.48%
Asian	All Respondents	68,309	2,397	303	2,700	3.55%	0.45%	4.01%
Black or African American	All Respondents	68,309	2,413	46	2,459	3.55%	0.05%	3.60%
American Indian/Alaska Native	All Respondents	68,309	2,472	29	2,501	3.52%	0.02%	3.54%
White	All Respondents	68,309	2,481	163	2,644	3.59%	0.22%	3.81%
Detailed Race: 15 Levels	All Respondents	68,309	2,520	499	3,019	3.73%	0.70%	4.43%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.26 Percentage of Cases Imputed for Income Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Amount of Income								
Respondent's Total Income > or < $20,000	All Respondents	68,309	822	31	853	1.85%	0.04%	1.89%
Respondent's Total Income (Finer Categories)	All Respondents	68,309	1,833	0	1,833	4.56%	0.00%	4.56%
Total Family Income > or < $20,000	All Respondents	68,309	2,985	0	2,985	3.96%	0.00%	3.96%
Total Family Income (Finer Categories)	All Respondents	68,309	6,881	1,272	8,153	9.96%	4.93%	14.88%
Source of Income								
Family Received Income from Job	All Respondents	68,309	205	0	205	0.22%	0.00%	0.22%
Family Received Social Security or Railroad Retirement Payments	All Respondents	68,309	711	0	711	0.60%	0.00%	0.60%
Family Received Public Assistance	All Respondents	68,309	531	0	531	0.44%	0.00%	0.44%
Family Received Supplemental Security Income	All Respondents	68,309	975	0	975	0.83%	0.00%	0.83%
Respondent/Other Family Member Received Food Stamps	All Respondents	68,309	327	0	327	0.34%	0.00%	0.34%
Family Received Welfare/Job Placement/Child Care	All Respondents	68,309	401	0	401	0.31%	0.00%	0.31%
Number of Months on Welfare	Family Receives Public Assistance or Welfare/Job Placement/Child Care	4,833	222	0	222	4.08%	0.00%	4.08%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.27 Percentage of Cases Imputed for Health Insurance Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Respondent Has Health Insurance								
As Defined by the 1999 Survey Method	All Respondents	68,309	596	2	598	0.41%	0.01%	0.41%
As Defined by the 2001 Survey Method	All Respondents	68,309	614	2	616	0.43%	0.01%	0.44%
As Defined by the Constituent Variables Method	All Respondents	68,309	614	0	614	0.43%	0.00%	0.43%
Type of Insurance								
Private – Consistent with Pre-1999 Surveys	All Respondents	68,309	519	0	519	0.40%	0.00%	0.40%
Private – As Defined by the Constituent Variables Method	All Respondents	68,309	519	0	519	0.40%	0.00%	0.40%
Medicare	All Respondents	68,309	264	0	264	0.21%	0.00%	0.21%
Military Health Care (CHAMPUS, TRICARE, CHAMPVA, VA)	All Respondents	68,309	285	0	285	0.21%	0.00%	0.21%
Medicaid/CHIP	All Respondents	68,309	660	0	660	0.55%	0.00%	0.55%
Other	Respondent Does Not Have Private Health Insurance, Medicare, Medicaid/CHIP, or CHAMPUS	12,025	208	0	208	0.94%	0.00%	0.94%

CHIP = Children's Health Insurance Program; CHAMPUS = Civilian Health and Medical Program of the Uniformed Services; CHAMPVA = Civilian Health and Medical Program of the Department of Veteran's Affairs; VA = Department of Veteran's Affairs.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.28 Percentage of Cases Imputed for Roster Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Household Roster								
Number of Persons in Household	All Respondents	68,309	42	0	42	0.06%	0.00%	0.06%
Number of Children in Household Aged < 18	All Respondents	68,309	225	0	225	0.29%	0.00%	0.29%
Number of Persons in Household Aged ≥ 65	All Respondents	68,309	377	54	431	0.34%	0.06%	0.40%
Family Roster								
Presence of Family Members in Household	All Respondents	68,309	53	0	53	0.07%	0.00%	0.07%
Number of Respondent's Family Members in Household excluding Foster Relationships	All Respondents	68,309	68	10	78	0.08%	0.01%	0.09%
Number of Respondent's Family Members in Household including Foster Relationships	All Respondents	68,309	75	3	78	0.09%	0.01%	0.09%
Number of Respondent's Family Members in Household Aged < 18 excluding Foster Relationships	All Respondents	68,309	152	0	152	0.19%	0.00%	0.19%
Number of Respondent's Family Members in Household Aged < 18 including Foster Relationships	All Respondents	68,309	158	0	158	0.20%	0.00%	0.20%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table A.29 Percentage of Cases Imputed for Pair Variables

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Pair Relationships								
Family Pair Relationship Indicator	All Pair Members	38,918	438	0	438	1.41%	0.00%	1.41%
Multiplicities								
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	3,962	2	0	2	0.03%	0.00%	0.03%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	7,496	6	0	6	0.14%	0.00%	0.14%
Multiplicity Count: Child-Parent, Parent Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	8,886	16	0	16	0.18%	0.00%	0.18%
Multiplicity Count: Child-Parent, Parent Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	3,534	4	0	4	0.24%	0.00%	0.24%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-14	Parent-Child Pair, Child is 12-14	3,962	96	0	96	2.57%	0.00%	2.57%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-17	Parent-Child Pair, Child is 12-17	7,496	182	0	182	2.62%	0.00%	2.62%
Multiplicity Count: Child-Parent, Child Focus, Child is 12-20	Parent-Child Pair, Child is 12-20	8,886	226	0	226	2.62%	0.00%	2.62%
Multiplicity Count: Child-Parent, Child Focus, Child is 15-17	Parent-Child Pair, Child is 15-17	3,534	86	0	86	2.68%	0.00%	2.68%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 12-14 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	4,382	22	0	22	0.49%	0.00%	0.49%
Multiplicity Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	Sibling-Sibling Pair: Younger is 12-14, Older is 15-17	4,382	24	0	24	0.54%	0.00%	0.54%

Table A.29 Percentage of Cases Imputed for Pair Variables (continued)

	Domain	Unweighted Frequencies				Weighted Percentages		
		Respondents in Domain	Imputed	Logically Assigned	Imputed or Logically Assigned	Imputed	Logically Assigned	Imputed or Logically Assigned
Multiplicity Count: Sibling-Sibling (12-17/18-25), 12-17 Focus	Sibling-Sibling Pair: Younger is 12-17, Older is 18-25	4,954	50	0	50	0.92%	0.00%	0.92%
Multiplicity Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	Sibling-Sibling Pair: Younger is 12-17, Older is 18-25	4,954	50	0	50	0.81%	0.00%	0.81%
Household Person-Level Count								
Household Count: Number of Spouse-Spouse Pairs in Household	All Respondents	68,309	118	0	118	0.31%	0.00%	0.31%
Household Count: Number of Spouse-Spouse Pairs with Children	All Respondents	68,309	69	0	69	0.13%	0.00%	0.13%
Household Count: Child-Parent, Parent Focus, Child is 12-14	All Respondents	68,309	281	0	281	0.35%	0.00%	0.35%
Household Count: Child-Parent, Parent Focus, Child is 12-17	All Respondents	68,309	436	0	436	0.59%	0.00%	0.59%
Household Count: Child-Parent, Parent Focus, Child is 12-20	All Respondents	68,309	496	0	496	0.65%	0.00%	0.65%
Household Count: Child-Parent, Child Focus, Child is 12-14	All Respondents	68,309	80	0	80	0.10%	0.00%	0.10%
Household Count: Child-Parent, Child Focus, Child is 12-17	All Respondents	68,309	151	0	151	0.20%	0.00%	0.20%
Household Count: Child-Parent, Child Focus, Child is 12-20	All Respondents	68,309	213	0	213	0.22%	0.00%	0.22%
Household Count: Sibling-Sibling (12-14/15-17), 15-17 Focus	All Respondents	68,309	50	0	50	0.06%	0.00%	0.06%
Household Count: Sibling-Sibling (12-17/18-25), 18-25 Focus	All Respondents	68,309	81	0	81	0.15%	0.00%	0.15%

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Appendix B: Detailed Documentation of Edits for Variables from the Noncore Demographics, Core Substance Use, and Noncore Self-Administered Sections in the 2012 NSDUH

Appendix B: Detailed Documentation of Edits for Variables from the Noncore Demographics, Core Substance Use, and Noncore Self-Administered Sections in the 2012 NSDUH

B.1 Introduction

This appendix provides detailed documentation of edits that were implemented (where applicable) for variables from the noncore demographics, core substance use, and noncore self-administered sections. The remaining sections of the appendix are organized as follows:

- Section B.2 presents tables describing edits for selected noncore demographics variables corresponding to editing procedures that were discussed in Chapter 5.
- Section B.3 presents tables describing edits for core substance use variables corresponding to editing procedures that were discussed in Chapter 6.
- Section B.4 presents tables describing edits for variables in noncore self-administered modules corresponding to editing procedures that were discussed in Chapter 7.

The issues that are shown in the tables in these sections reflect cumulative response patterns that have been observed in the data since a given module or set of questions was added to the interview. However, not all of these patterns may be encountered in a given survey year. In addition, no editing was done for some types of responses that are presented in the tables. Nevertheless, this information is included to document the types of noteworthy responses for which the data were not edited.

B.2 Detailed Documentation of Edits for Selected Noncore Demographics Variables

This section presents tables describing edits for the noncore education and the employment and workplace variables from the noncore demographics.

- Table B.1 presents edits pertaining to variables in the noncore education section.
- Table B.2 presents edits pertaining to variables in the employment and workplace section.

Table B.1 Edits Pertaining to the Noncore Education Section

Issue	Edits Implemented
The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was still in school (LFSCHWHY = 778).	The R was logically inferred to be currently enrolled in school. A special code of 3 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: HSDIPLMA (receipt of a high school diploma), HSGED (receipt of a GED certificate), and LFTSCHAG (age when the R left school). Consequently, any data in these items were overwritten in the edited variables with codes to indicate logical inference that these questions should have been skipped (see Section 2.4.2). Data were not overwritten for LFSCHWHY (reason for leaving school) because that was the variable responsible for inferring that the R was currently enrolled.
The R reported not being currently enrolled in school. In the question about reasons for leaving school without getting a high school diploma (LFSCHWHY), however, the R specified that he or she was being homeschooled (LFSCHWHY = 601).	The R was logically inferred to be currently enrolled in school. A special code of 5 was assigned to the edited school enrollment variable SCHENRL. As above, any data in HSDIPLMA, HSGED, and LFTSCHAG were overwritten. Data were not overwritten for LFSCHWHY because that was the variable responsible for inferring that the R was currently enrolled. If the R was aged 12 to 18, a code of 14 was assigned to the variable SCHTYPE (Homeschooled, LOGICALLY ASSIGNED) to indicate that the R was logically inferred to be homeschooled. No editing was done to SCHTYPE if the R was older than 18. In addition, this edit was not implemented if interviewers specified that the respondent had *completed* homeschooling. In these latter situations, LFSCHWHY was assigned a code other than 601 (see below).
The R reported being currently enrolled in school, and a hard error was triggered between QD11 and QD18. The FI's comments for suppressing the hard error indicated that the R was currently enrolled in technical or vocational school.	The R was logically inferred not to be currently enrolled in school. A special code of 4 was assigned to the edited school enrollment variable SCHENRL. For the following variables, it was logically inferred that they should have skipped: EDUCATND (current grade), SDNTFTPT (full- or part-time status), SCHDSICK (number of days in the past 30 days that the R missed school because the R was sick), and SCHDSKIP (number of days the R skipped school in the past 30 days). Consequently, any data in these items were overwritten in the edited variables with codes to indicate that these items should have been skipped.
The R did not know or refused to report whether he or she was enrolled in school, reported being on a holiday or break from school (QD17A = 1), but reported that he or she did not intend to return to school once the break was over (QD17B = 2).	The R was defined as not being enrolled in school (SCHENRL=2). The variables pertaining to the current grade through the number of days that the R skipped school in the past 30 days (EDUCATND, SDNTFTPT, SCHDSICK, and SCHDSKIP) were assigned legitimate skip codes.
The R was aged 23 or younger and reported not being enrolled in school. However, the interview was conducted in June, July, or August (i.e., when school was not in session). The R also originally reported getting a high school diploma but was inferred not to have received one (i.e., the R completed the 9th grade or lower). In addition, the highest completed grade was within 1 year (in either direction) of the expected completed grade for that R's age.	A code of 52 was assigned to the school enrollment variable SCHENRL only if there was some ambiguity (i.e., a response of "don't know" or "refused") in the follow-up questions QD17B or QD17C, regarding the R being on break from school or intending to return to school when the R's break was over, respectively. This code of 52 was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status due to (a) the interview being conducted in summer months when most schools are not in session, and (b) ambiguity in the R's answers to QD17B or QD17C. Otherwise, no editing was done to SCHENRL (i.e., the R continued to be classified as not enrolled in school) when the R indicated in QD17B that he or she was not on break from or the R was on break but indicated in QD17C that he or she did not intend to return to school once this break was over.

Table B.1 Edits Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
The current grade (QD18) was potentially inconsistent with the highest grade that the respondent (R) reported completing (QD11), and (a) the hard error between QD11 and QD18 was not triggered (e.g., the current grade from QD18 was two or more grades higher than the highest grade from QD11); or (b) the hard error was triggered and suppressed, but the field interviewer (FI) did not provide sufficient information to determine what corrections needed to be made.	An algorithm was developed that compared the self-reported current and highest grades with the R's current age (see text). A noncore edited variable for the highest grade completed (EDTEDUC) also was created. Edits were generally implemented as follows: • When both the current grade and the highest completed grade were potentially consistent with the R's age, the edits picked the response from QD18 or QD11 that would yield the most consistent data. The second variable in the pair then was edited for consistency with the response that was picked as being most consistent. • When the current grade was more consistent with the R's current age than was the reported highest grade from the core demographics, then EDTEDUC was logically assigned a code to indicate that the R had completed the lower grade that was adjacent to his or her current grade. • When the highest completed grade was more consistent with the R's current age than was the reported current grade, then the edited current grade (EDUCATND) was logically assigned a code to indicate that the R was in the next highest grade relative to the one he or she had completed, or else EDUCATND was coded as bad data. When neither the current grade (QD18) nor the highest completed grade (QD11) was consistent with the R's age, either EDTEDUC or EDUCATND (or both) were coded as bad data. If the current grade was exactly two grades higher than the last grade but the highest grade was lower than the expected highest grade, then EDTEDUC was coded as bad data. If the current grade was more than two grades higher than the last grade but the current grade was lower than the expected current grade, then EDUCATND was coded as bad data. If the current grade was lower than the highest grade, the one that was closest to the expected grade was chosen, and the other was set to bad data. If both EDTEDUC and EDUCATND were both close to their expected grades, both were set to bad data.
The R reported not being currently enrolled in school, reported receiving a high school diploma, but reported completing the 10th or 11th grade.	No editing was done, and the variable pertaining to receipt of a high school diploma (HSDIPLMA) retained a value of 1 (i.e., "yes"). The rationale was that the R may have gone through school on an accelerated pace or may have otherwise qualified for a high school diploma with fewer than 12 years of education (e.g., if the R went to school in another country).
The R reported not being enrolled in school but having received a high school diploma. However, the R had completed only the 9th grade or lower.	The R was logically inferred in HSDIPLMA not to have received a high school diploma. HSDIPLMA was assigned a code of 4 (No LOGICALLY ASSIGNED).
The R reported not being enrolled and not having received a high school diploma. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had gotten a diploma. That included situations where the R may have received a diploma in another country.	The R was logically inferred to have received a high school diploma, provided that the R had completed the 10th grade or higher. The edited variable HSDIPLMA was assigned a code of 3 (Yes LOGICALLY ASSIGNED). Although Rs would have skipped the question about getting a GED if they had answered "yes" to the question about getting a high school diploma (QD22), no editing was done to the GED variable HSGED (QD23) when Rs were logically inferred to have gotten a high school diploma. In addition, no editing was done to LFSCHWHY and LFTSCHAG in this situation.

Table B.1 Edits Pertaining to the Noncore Education Section (continued)

Issue	Edits Implemented
The R reported not being enrolled in school, not having received a high school diploma, and not having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had received a GED (LFSCHWHY = 606).	The R was logically inferred to have gotten a GED certificate. For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received a GED, not that he or she was working on a GED.
The R reported not being enrolled in school, not having received a high school diploma, and having received a GED certificate. In the question about reasons for leaving school without getting a high school diploma, however, the R specified that he or she had completed homeschooling but had not received a GED (LFSCHWHY = 621).	The R's response of "yes" in HSGED was considered to be questionable based on the response in LFSCHWHY. HSGED was set to a value of 11 (i.e., "bumped" by 10). For this edit to be implemented, however, the R had to have indicated explicitly that he or she had actually received no GED after having completed homeschooling.
The R reported leaving school at an age greater than his or her current age.	The edited variable corresponding to the age at leaving school was assigned a bad data code.
The R reported leaving school at age 3 or younger, or the R reported leaving school at an age that was considered too young for the highest grade that he or she reported completing (e.g., completed the 11th grade but reported leaving school at age 13 or younger).	The edited variable corresponding to the age at leaving school was assigned a bad data code.
The number of days in the past 30 that the R missed school because the R was sick and the number of days that the R skipped school added up to more than 30.	No editing was done in this situation.
The R reported being currently enrolled in school but reported skipping school all 30 days in the past 30 days.	A code of 11 was assigned to the school enrollment variable SCHENRL. This code was intended to indicate to analysts that there was some uncertainty about the R's current enrollment status.
The R reported being currently enrolled in school but reported in question QD20 that he or she missed school because of sickness for more than 30 days. This pattern was observed in earlier years because a code of 90 was used to mean "school not in session," and the computer-assisted interviewing (CAI) program code did not allow for discontinuities in the allowable range.	Values of 31 days were set to 30 days. Values greater than 31 days but fewer than 90 (i.e., school not in session) were replaced with bad data codes. This logic was in place, but values greater than 31 and fewer than 90 did not occur in the data.

Table B.2 Edits Pertaining to the Employment and Workplace Section

Issue	Edits Implemented
The R did not know or refused to report in question QD26 whether he or she worked in the past year. However, the R also reported in question QD33 (edited variable WRKEDYR) that he or she did not have a job in the past 12 months.	The R was logically inferred not to have worked in the past week (WRKEDWK = 4) and not to have had a job in the past week (WRKHAVJB = 4), where 4 = No LOGICALLY ASSIGNED. Subsequent employment and workplace variables that could be assigned legitimate skip codes were edited as though QD26 and QD27 had been answered as "no."
The R reported working in the past week. However, the R subsequently reported missing work for all 30 of the past 30 days because he or she was sick or did not want to be at work (or both). Because the past week was included in the 30 days prior to the interview, it would be inconsistent for an R to report working in the past week but missing work for every day in the past month.	The following edits were implemented in this situation: • If the R reported that he or she missed work for all 30 days in the past month because he or she was sick, the edited variable (WORKDAYS) was assigned a bad data code. • If the R reported missing work for all 30 days in the past month because he or she did not want to be there, the edited variable (WORKBLAH) was assigned a bad data code.
The respondent (R) reported working in the past week in question QD26. However, the R subsequently reported being without a job at some point in the past 12 months and reported being without a job during all 52 weeks in the past 12 months. Because all 52 weeks of the 12-month period prior to the interview would include the week prior to the interview, it would be inconsistent for an R to report working in the past week but not working for all 52 weeks in the past year.	The edited variable pertaining to the number of weeks without a job in the past 12 months (WRKUNWKS) was assigned a bad data code.
The R answered question QD26 (worked in the past week) as "no" but answered question QD27 (having a job in the past week) as "don't know." Edit logic prior to 2003 left the variable pertaining to the number of hours worked in the past week WRKHRSWK as blank (i.e., a legitimate skip code was not assigned to WRKHRSWK). The prior logic for assigning legitimate skip codes to WRKHRSWK was part of the logic for assigning legitimate skips when both QD26 and QD27 were answered as "no." However, only the response to question QD26 truly applies to WRKHRSWK.	Since 2003, a legitimate skip code has been assigned to WRKHRSWK when QD26 is answered as "no" (QD26 = 2), independent of how QD27 was answered.
The reported year when the R last worked for pay was fewer than 5 years from the R's birth year (including situations where the year the R reported last working for pay was earlier than the year the R was born).	The edited variables WRKLSTMN and WRKLSTYR were assigned bad data codes.

Table B.2 Edits Pertaining to the Employment and Workplace Section (continued)

Issue	Edits Implemented
The R was not asked whether he or she was self-employed in the past 12 months because the R had already given an answer indicating that he or she had been self-employed. This could occur in one of the ways listed below. • The R reported not working in the past week because he or she was self-employed and did not have any business (QD30 = 5). • The R reported in question INOC06 that the category that best described the business in which he or she worked was one in which the R was self-employed (INOC06 answered as 7 or 8).	The edited variable pertaining to self-employment in the past 12 months (WRKSLFEM) was assigned a code to indicate that "yes" could be logically inferred. This was done instead of assigning a legitimate skip code. This edit did not apply if INOC06 indicated that Rs worked without pay in a family business or farm.
The R did not report being self-employed at any time in the past 12 months but reported having a job. However, the industry and occupation (I&O) question pertaining to the R's last job (INOC08) indicated that the R was self-employed in an incorporated or unincorporated business (edited variable WRKBZCYR, corresponding to INOC08, had a value of 7 or 8).	The edited variable WRKSLFEM was logically inferred to have been answered as "yes," provided that the following conditions held: • The R reported working in the past year (WRKEDYR = 1), such that reported self-employment in INOC08 would pertain to self-employment in the past year. • The year and month that the R reported last working for pay (WRKLSTYR and WRKLSTMN) also were consistent with the R reporting that he or she worked in the past year. The following data in WRKLSTYR and WRKLSTMN were considered to be consistent (or at least not contradictory) with indications that the R worked in the past year (WRKEDYR = 1): • The R reported last working for pay in the current interview year. • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was within 12 months of the interview date, or was the same month as the interview date. • The R reported last working for pay in the previous year, but the month that the R reported last working for pay had a missing value. In this situation, WRKEDYR = 1, and an indication of self-employment in INOC08 was still allowed to infer in WRKSLFEM that the R had been self-employed in the past 12 months. WRKSLFEM was not logically inferred to be "yes" if the R reported working in the past year (WRKEDYR = 1), WRKBZCYR = 7 or 8, but any of the following occurred: • The R reported last working for pay in the previous year, and the month that the R reported last working for pay was more than 12 months beyond the interview date. • The R had missing data for the year when he or she last worked for pay (e.g., if WRKLSTYR was refused). • A problem had been identified with the interview date that was stored by the computer-assisted interviewing (CAI) system while the interview was in progress.

B.3 Detailed Documentation of Edits for Core Substance Use Variables

This section presents tables that detail the edits that were applied to variables in the core substance use modules. Because data on use of methamphetamine and other stimulants from the noncore special drugs module were used to create core-plus-noncore (CPN) recency variables for these substances as described in Section 6.2.6, Section B.3 also includes details about the editing of variables that were relevant to creating these CPN recency variables.

- Tables B.3 and B.4 present edits pertaining to variables for lifetime substance use, corresponding to the discussion of procedures in Section 6.2.1.
- Tables B.5 to B.7 present edits pertaining to variables for the frequency of use of different substances in the past 12 months or past 30 days, corresponding to the discussion of procedures in Section 6.2.4. These tables also document procedures for editing the frequency-of-use variables for "parent/child" drug combinations (e.g., pain relievers and OxyContin®) when the only substance that the respondent used was the child drug (Section 6.2.3).
- Tables B.8 to B.10 present edits pertaining to variables for incidence (i.e., initiation of use), corresponding to the discussion of editing procedures in Section 6.2.5. These tables also document procedures for editing the incidence variables for "parent/child" drug combinations when the only substance that the respondent used was the child drug (Section 6.2.3).
- Table B.11 presents edits for variables on methamphetamine prevalence, stimulant prevalence, and related follow-up variables in the noncore special drugs, corresponding to the discussion of procedures in Section 6.2.6 for creating the CPN recency variables for these substances.
- Finally in this section, Table B.12 presents miscellaneous edits pertaining to the tobacco variables.

Table B.3 Edits Pertaining to Lifetime Variables for Hallucinogens through Sedatives Based on "OTHER, Specify" Data

Issue	Edits Implemented
Respondents (Rs) could specify something that corresponded to a drug in the list. For example, if an R specified use of "marijuana laced with PCP" as "some other hallucinogen," this response would indicate PCP use—even if such use had not previously been reported.	If the corresponding drug in the list was not answered as "yes," then "yes" was logically inferred, and the "specify" response was retained to indicate to analysts the source of the logical inference.
Rs could use street names or slang terms when specifying their use of some other drug besides the ones they previously had been asked about. Use of these slang terms to infer use of a drug that Rs had already been asked about (e.g., LSD, PCP) could be problematic, however, because the meaning of a particular slang term could vary by region, could apply to more than one drug, or could change over time.	Use of street or slang terms to infer lifetime use of a specific drug was generally restricted to situations where that term was supplied to respondents as a synonym for that drug. For example, "angel dust" was listed in question LS01b as an alternative term for PCP. Thus, situations where Rs specified angel dust were used to infer lifetime use of PCP, if question LS01b had not already been answered as "yes." However, other potential slang terms for PCP (e.g., "boat") that were not listed in question LS01b were not used in editing. Additional situations where slang terms were classified with particular drugs included "shrooms" (coded as psilocybin/mushrooms), "X" and "X-TC" (coded as Ecstasy), "Roofies" (coded as Rohypnol®), and "rock" (coded as crack cocaine). In addition, question ST01 in the stimulants module listed "speed" as an alternative term for methamphetamine. However, the decision was made that indications of speed in and of themselves should not be used to infer lifetime methamphetamine use because speed may often be used to refer to other stimulants or to stimulants as a group, not just methamphetamine.
Rs could specify use of some other drug within a particular category of drugs, but the drug being mentioned applies to another drug category covered in the survey (e.g., specifying Valium®, a tranquilizer, when asked about nonmedical use of pain relievers).	No editing across modules was done when these types of responses occurred (Section 2.4.4). However, the "OTHER, Specify" data within a module indicate to analysts when Rs have specified use of a drug that fits another category. This approach assumed that some Rs may specify the use of other drugs according to their *functional* properties. For example, Rs may specify certain tranquilizers in the sedatives section because tranquilizers can cause drowsiness. Similarly, the definition of hallucinogens (i.e., drugs that "often cause people to see or experience things that are not real") could apply to other drugs besides hallucinogens that alter one's perception of reality.

Table B.4 Edits Pertaining to Reports of Over-the-Counter Drugs in "OTHER, Specify" Data for Pain Relievers through Sedatives

Issue	Edits Implemented
Rs may indicate lifetime use of one or more prescription-type psychotherapeutics (e.g., Darvon®, Darvocet-N®, or Tylenol® with codeine) but specify only over-the-counter (OTC) medications as the other drug that they used (e.g., specifying only Tylenol® as the other prescription pain reliever that they used nonmedically).	Because Rs were instructed in the psychotherapeutics section not to include use of OTC medications, edits in this situation inferred that Rs logically had not used some other drug in that category. However, Rs who reported lifetime nonmedical use of a prescription-type medication in response to a previous question were still at least lifetime abusers of that category of drugs. This edit also was done in situations where Rs answered at least one prescription-type gate question as "don't know" (DK) or "refused" (REF). This edit was *not* done if Rs reported use of "some other" psychotherapeutic drug and (a) specified use of a prescription-type medication and an OTC drug, or (b) had some response of DK or REF along with the indication of OTC use. In this latter situation, the assumption was that a response of DK or REF meant that the R was still potentially a nonmedical user of a prescription-type drug.
The *only* indication of lifetime nonmedical use was an OTC drug that was specified. (In this situation, unlike the one described above, Rs have denied ever using other prescription-type medications that were covered in the particular category.)	Because Rs were instructed in the psychotherapeutics sections not to include use of OTC medications, edits in this situation inferred that the R had never used prescription or street psychotherapeutics. This edit was implemented if only OTC drugs were specified, or the data included only OTC drugs and codes for "bad data." This edit was not done in situations where one or more other gate questions for a type of psychotherapeutic medication were answered as DK or REF because that was not conclusive evidence that the R had *never* used. This edit also was not done in situations where the "OTHER, Specify" data included responses of DK or REF in addition to reports of OTC drugs (i.e., but no reports of nonmedical use of prescription-type drug). As for the previous issue, responses of DK or REF were interpreted to mean that the R was still potentially a nonmedical user of a prescription-type drug.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use

Issue	Edits Implemented
Number of days that the respondent (R) used in the past 12 months is greater than 365 days (e.g., if the R reported using 31 days per month on average, then 31 x 12 = 372 days).	Nonmissing values greater than 365 were trimmed back to a maximum of 365 days. A flag was set to indicate that this trimming was done.
Number of days that the R reported using in the past 30 days implies less use in the past 12 months than what the R reported in the 12-month frequency (e.g., if the R reported using on only 1 day in the past 30 days and did not use on 29 days, the maximum number of days that the R could have used in the past 12 months would be 336 days); this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The 12-month frequency was trimmed to be consistent with the number of days that the R reported using in the past 30 days. The relevant flag variable indicated that this trimming was done.
Rs who answered the 30-day frequency questions as "don't know" (DK) or "refused" (REF) had the opportunity to give their best estimate of the number of days they used in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). For alcohol through inhalants, the maximum possible value of the "estimated" 30-day frequency could imply less use in the past 12 months than what the R reported in the 12-month frequency. For example, if the R reported using a drug on "10 to 19 days" in the past 30 days, the R would not have used the drug on at least 11 days in that period. It would therefore be inconsistent for the R to have reported using the drug on more than 354 days in the past 12 months.	The 12-month frequency was trimmed to be consistent with the minimum number of days that the R estimated using in the past 30 days. For example, if the R estimated using a drug on "10 to 19" days in the past 30 days, it was assumed that the R could have used on 345 days in the past 12 months, which would be consistent with use on 10 days in the past 30 days (and nonuse on 20 days). This value would not be considerably different from that derived by assuming use on the maximum number of estimated days in the past 30 days (e.g., 354 possible days of use in the past 12 months, if the R reported using on 10 to 19 days in the past 30 days).
For Rs who estimated their 30-day frequency, the final 12-month frequency was consistent with some, but not all, values in the range for the estimated 30-day frequency.	No editing was done to the 12-month frequency. For example, if an R reported using a drug on "6 to 9 days" in the past 30 days and reported using on 8 days in the past 12 months, the 12-month frequency was within the range of 6 to 8 days. Thus, a 12-month frequency of 8 days would be consistent with the estimated 30-day frequency, as long as the R used on 6 to 8 days in the past 30 days (and did not use on 9 days in that period). In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example, the allowable range for assigning the final 30-day frequency was 6 to 8 days rather than 6 to 9 days.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
R initiated use at some point in the past 12 months. Relative to when the R was interviewed, the R therefore could not have used over the entire 12-month period.	The value of the 12-month frequency was reduced (i.e., prorated) according to the maximum possible allowable number of days that the R could have used in the past 12 months, relative to when the R was interviewed. If the R answered the 12-month frequency in terms of a total number of days used in that period and the answer was greater than the maximum number of days that the R could have used, the 12-month frequency was reduced to this upper limit. The maximum possible allowable number of days was determined from the interview date and the month- and year-of-first use or the date of the R's last birthday, whichever yielded the smallest number of days that the R could have used. If the R reported first using in the survey year but did not report the month when he or she first used, it was assumed that the R potentially started use in January of the survey year.
The R initiated use in the past 12 months but definitely initiated use more than 30 days prior to the interview date. The R also reported use on only 1 day in the past 12 months. However, the recency had a value of 1 (Used in the past 30 days) or 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED).	The 12-month frequency of 1 day was set to bad data. If the R last used in the past 30 days and the R initiated use more than 30 days ago but within the past 12 months, the R logically would have used on at least 1 day in the past 30 days and on at least 1 other day in the period that was more than 30 days ago but within the past 12 months. Similarly, a recency of 8 indicated that there was uncertainty about whether the R last used in the past 30 days or more than 30 days ago but within the past 12 months. Consequently, use on only 1 day in the past 12 months and initiation more than 30 days ago within the past 12 months would have forced the recency to be more than 30 days ago. In addition, when the recency had indicated use in the past 30 days and the above edit had changed a 12-month frequency from 1 day to bad data, whatever values that existed in the 30-day variables were retained. In particular, if the 30-day frequency had a missing value, that missing value was retained. (Other edits described below would have logically inferred that the R used on 1 day in the past 30 days if the R also had indicated use on exactly 1 day in the past 12 months.)
Information was needed for establishing constraints on donors for the imputation of the 12-month frequency-of-use variables. For example, if a 12-month frequency variable had been changed from 1 day to bad data, the imputation team needed to constrain the imputation of the 12-month frequency to a value greater than 1 day.	Indicator flags were created to aid the statistical imputation team in establishing specific constraints for the imputation of the 12-month frequency-of-use variables. The indicator flags had the following values: 1 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN 1 DAY. 2 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF USE. (Did not apply to the psychotherapeutics, which do not have 30-day frequency variables.) 3 = REQUIRE DONOR WITH 12-MONTH FREQUENCY OF USE GREATER THAN RECIPIENT'S 30-DAY FREQUENCY OF BINGE ALCOHOL USE. (Applied only to alcohol).

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
R initiated use at some point in the past 30 days. Relative to when the R was interviewed, however, the 30-day frequency (or the minimum value of the "estimated" 30-day frequency) exceeded the maximum possible number of days that the R could have used the drug in that period.	If the R gave an exact number of days for the 30-day frequency, the value was reduced to agree with the maximum possible number of days that the R could have used the drug. If the R estimated the 30-day frequency as a range of a possible number of days, the range for the estimated 30-day frequency was reduced to agree with the maximum possible number of days that the R could have used. For example, if the R estimated using a drug on 20 to 29 days in the past 30 days but could have used the drug at most on 15 days in that period, the estimated 30-day frequency was revised to the range indicating use on 10 to 19 days.
The 30-day frequency was greater than 12-month frequency; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The value for the 30-day frequency was retained, and the 12-month frequency was set to a missing value (i.e., bad data). (Any of the corresponding 12-month frequency source variables that the respondent answered were set to bad data as well.) This edit conserved data reported by the R that suggest more frequent use in the past 30 days. Although both the 12-month and 30-day frequencies could have been set to missing (i.e., bad data) and subsequently imputed, a drawback of this approach is that a donor could be chosen whose 30-day and 12-month frequency data were consistent but whose data were considerably different from what the recipient R originally reported, especially if the recipient R's original answer to the 30-day frequency was at or close to 30.
For cigarettes, chewing tobacco, snuff, and cigars, Rs who answered the 30-day frequency question as DK or REF were asked to give their best estimate of the number of days they used these tobacco products in the past 30 days. These best estimate variables were categorical as opposed to being continuous (e.g., 1 = 1 or 2 days, 2 = 3 to 5 days). Because there were no 12-month frequency questions for tobacco products, inconsistencies would not exist between these "estimated" 30-day frequency variables and other data for these tobacco products.	No editing was done to the corresponding 30-day frequency variables (i.e., leaving the original answer of DK or REF). A final value was assigned by the statistical imputation team.
For alcohol through inhalants, if Rs gave their best estimate of the number of days they used in the past 30 days, it was possible for the bottom end of the range for an "estimated" 30-day frequency to exceed the value for the 12-month frequency (i.e., after any opportunities that the R had for inconsistency resolution).	The value indicating the range for the "estimated" 30-day frequency was retained, and the 12-month frequency was set to a missing value (i.e., bad data). For example, if the R reported using a drug on 5 days in the past 12 months but estimated use of that drug on 6 to 9 days in the past 30 days, there would be no way for these two answers to be consistent. Setting the 12-month frequency to bad data in this situation is consistent with the edit when Rs report an exact number of days for the final 30-day frequency that is greater than the final 12-month frequency.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
The bottom end of the range for the final "estimated" number of days that the R used a drug in the past 30 days is *exactly* equal to the final 12-month frequency; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The final, edited 30-day frequency was set to the value for the 12-month frequency. The code was retained for the "estimated" 30-day frequency to indicate to analysts where the value came from for the 30-day frequency. For example, if the R reported using a drug on exactly 6 days in the past 12 months and estimated using that drug on 6 to 9 days in the past 30 days, the 30-day frequency was set to exactly 6 days. The two answers in this example are not necessarily inconsistent, as long as the R used on exactly 6 days in the past month. Thus, the response of 6 days in the past 12 months could be thought of as "trapping" the possible number of days that the R could have used in the past 30 days.
The 30-day frequency has a final response of DK or REF (i.e., the R did not give a best estimate for the 30-day frequency), but the 12-month frequency indicates use on exactly 1 day.	The final, edited 30-day frequency was logically inferred to be 1 day. If the R last used in the past 30 days and used on only 1 day in the past 12 months, that 1 day of use had to have occurred in the 30 days prior to the interview.
The value for the final 12-month frequency is somewhere within the range for the final "estimated" number of days that the R used a drug in the past 30 days; this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	No further editing was done because the data were not necessarily inconsistent. For example, if the R reported using a drug on 8 days in the past 12 months and estimated use of that drug on 6 to 9 days in the past 30 days, use of the drug on 6, 7, or 8 days in the past 30 days would still be consistent with use on 8 days in the past 12 months. In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency. This procedure narrowed the allowable range of the 30-day frequency for consistency with the 12-month frequency, with the final 30-day frequency being picked from within that narrowed range. In the above example where the R reported using on 8 days in the past 12 months and 6 to 9 days in the past 30 days, the allowable range for the 30-day frequency would be 6 to 8 days, not 6 to 9, in order for the 30-day frequency to be consistent with the 12-month frequency.

Table B.5 Edits Pertaining to 12-Month and 30-Day Frequency Variables Other than for Parent/Child Relationships or Binge Alcohol Use (continued)

Issue	Edits Implemented
Overall 12-month frequency does not agree with the reported preference for giving the answer (e.g., if the R originally indicated that he or she would give an answer in terms of a total number of days per year, answered the days per year question as DK/REF, and then gave an answer in average days per month). This is a function of the computer-assisted interviewing (CAI) logic, which keeps routing Rs through the series until they give a nonmissing answer or answer all 12-month frequency questions as DK/REF; this applied to the alcohol through psychotherapeutics sections. Similarly, the R could change his or her preferred way of reporting the 12-month frequency in response to a consistency check with the 30-day frequency (e.g., starting out as reporting in terms of the total number of days in the past 12 months but changing to average number of days per month in response to a consistency check); this applied to alcohol through inhalants (but not to LSD, PCP, or Ecstasy).	The preferred way of reporting the 12-month frequency that best agreed with the final answer to the 12-month frequency was assigned to the "best way" variable (e.g., ALBSTWAY for alcohol), pertaining to the R's preferred way of answering the 12-month frequency questions. Thus, if an R started out as preferring to give an answer in total days per year but the final answer came from an average number of days per month, the preferred method of reporting would be consistent with the reporting in days per month. Any other 12-month frequency questions that the R answered that did not correspond with the final result were overwritten with LEGITIMATE SKIP Logically assigned codes. In the above example, the initial answer in terms of the total number of days used in the past 12 months would be overwritten, and the final answer of average number of days per month would be retained.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants

Issue	Edits Implemented
The 12-month frequency of methamphetamine use was greater than the corresponding frequency of any stimulant use. The same issue applied in the pain relievers module when the 12-month frequency of OxyContin® use was greater than the corresponding frequency of any pain reliever use.	The higher value from the methamphetamine frequency was assigned to the stimulant frequency. This also involved moving over the source variables from methamphetamine that were associated with the overall methamphetamine frequency. In addition, the methamphetamine variable associated with the preferred method of reporting the 12-month frequency (STBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the methamphetamine frequency variables. The same edits described above for methamphetamine and any stimulants were applied to the 12-month frequency variables for any pain relievers when the 12-month frequency of OxyContin® use was greater than the corresponding frequency of any pain reliever use.
The 30-day or 12-month frequency of crack use is greater than the corresponding frequency of any cocaine use.	The higher value from the crack frequency was assigned to the cocaine frequency. For the 12-month frequency, this also involved moving over the source variables from crack that were associated with the overall crack frequency. In addition, the cocaine variable associated with the preferred method of reporting the 12-month frequency (CCBSTWAY) was bumped by a value of 20 to indicate that data had been moved over from the crack cocaine variables.
Respondents (Rs) indicated that their month of first use for any cocaine, any pain relievers, or any stimulants was in the same month they were interviewed, and they used cocaine, pain relievers, or stimulants on only 1 day in the past 12 months. However, the Rs also were users of crack cocaine, OxyContin®, or methamphetamine, respectively, and they had missing data for the corresponding 12-month frequency variables pertaining to crack cocaine, OxyContin®, or methamphetamine.	Logically, these Rs also would have used crack cocaine, OxyContin®, or methamphetamine on 1 day in the past 12 months. Therefore, the corresponding 12-month frequency variables for crack cocaine, methamphetamine, or OxyContin® were set to 1 day in this situation.
Rs indicated that their month of first use for any cocaine, any pain relievers, or any stimulants was in the same month they were interviewed, and they used cocaine, pain relievers, or stimulants on only 1 day in the past 12 months. However, the Rs also were users of crack cocaine, OxyContin®, or methamphetamine, respectively, and they had missing data for the corresponding 12-month frequency variables pertaining to crack cocaine, OxyContin®, or methamphetamine.	Logically, these Rs also would have used crack cocaine, OxyContin®, or methamphetamine on 1 day in the past 12 months. Therefore, the corresponding 12-month frequency variables for crack cocaine, methamphetamine, or OxyContin® were set to 1 day in this situation.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants (continued)

Issue	Edits Implemented
In the stimulants module, the R was a lifetime user only of methamphetamine and reported last using methamphetamine in the past 30 days or more than 30 days ago but within the past 12 months. The same issue applied in the pain relievers module when the R was a lifetime user only of OxyContin® and reported last using OxyContin® at some point in the past 12 months.	Values from the methamphetamine 12-month frequency variables were assigned to the corresponding stimulant 12-month frequency variables. The variable STBSTWAY (preferred way of answering the stimulant 12-month frequency questions, corresponding to question ST10) was assigned a code in the 20s to indicate that data had been assigned from the methamphetamine 12-month frequency variables. Similarly, when the R was a user only of OxyContin®, values from the OxyContin® 12-month frequency variables were assigned to the corresponding pain reliever 12-month frequency variables, and the variable PRBSTWAY (preferred way of answering the pain reliever 12-month frequency questions) was assigned a code in the 20s.
The R was skipped out of the 12-month or 30-day frequency questions for cocaine because of his or her original answer to the cocaine recency, but nonmissing values existed in the corresponding frequency variables for crack. For example, if an R reported last using any cocaine more than 30 days ago but within the past 12 months, but reported last using crack within the past 30 days, 30-day frequency data could exist for crack but not for any cocaine.	No editing was done to the 12-month or 30-day cocaine frequency that had been skipped. In the imputation stage, the cocaine 12-month or 30-day frequency was subsequently edited or imputed for consistency with the corresponding crack frequency data.
The R estimated the frequency of use of both cocaine and crack in the past 30 days, with the estimated frequency for crack being greater than the estimated frequency for any cocaine.	The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST.
The "estimated" 30-day frequency of crack use is greater than the total number of days that the R reported using any cocaine in that same period.	The "estimated" value for the crack 30-day frequency (CK30EST) was assigned as the estimate for the number of days that the R used any cocaine in that period. The code assigned to the "estimated" cocaine 30-day frequency variable CC30EST was bumped by a value of 10 to indicate that data had been moved over from CK30EST. The edited cocaine 30-day frequency variable COCUS30A was assigned a bad data code to wipe out the R's original answer. For example, if the R reported using any cocaine on exactly 5 days in the past 30 days but reported using crack on 6 to 9 days in that period (i.e., a code of 3 in question CK06DKRE), the R was logically inferred to have used any cocaine on 6 to 9 days in the past 30 days. A code of 13 was assigned to CC30EST, corresponding to the code of 3 in CK30EST. In this example, it would be impossible for the R to have used crack on at least 6 days in the past 30 days but to have used any cocaine on only 5 days in that same period.

Table B.6 Edits Pertaining to Parent/Child 12-Month and 30-Day Frequency Variables for Cocaine, Pain Relievers, and Stimulants (continued)

Issue	Edits Implemented
The value for the cocaine 30-day frequency is somewhere within the range given by the "estimated" crack 30-day frequency.	Examination of preliminary data from 2000 indicated that all cocaine use in the past 30 days could not be automatically inferred to be crack use. Therefore, if the value for the cocaine 30-day frequency was the minimum possible number of days that the R could have used crack in that period, the crack 30-day frequency CRKUS30A was set to be equal to the cocaine 30-day frequency COCUS30A. For example, if the R reported using cocaine on exactly 6 days in the past 30 days and estimated using crack on 6 to 9 days in that period, it could be logically inferred that the R used crack on exactly 6 days. If the value for the cocaine 30-day frequency was greater than the minimum value for the estimated crack frequency, no further editing was done. However, information was provided to the statistical imputation team regarding the narrower range in which to assign a final value for the crack 30-day frequency. For example, if the R reported using cocaine on 7 days and crack on 6 to 9 days, the allowable range for the final crack 30-day frequency would be 6 or 7 days.

Table B.7 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use

Issue	Edits Implemented
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY, from AL08) is greater than the overall frequency of consumption of any alcohol in the past 30 days (ALCDAYS, from AL06), despite the presence of a consistency check.	If the respondent (R) reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days. If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from the five-drink frequency DR5DAY was retained and the overall 30-day frequency ALCDAYS was set to missing (i.e., bad data). This approach conserved data reported by the R that could indicate heavy alcohol use (i.e., consumption of five or more drinks on 5 or more days in the past 30 days).
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) is greater than the maximum possible value for the estimated frequency of consumption of alcohol in the past 30 days (AL30EST, from AL06DKRE).	If the R reported having five or more drinks per occasion on all 30 days, then the edits inferred that the R drank on all 30 days (i.e., the variable ALCDAYS was assigned a value of 30). This edit also was done if the R reported having five or more drinks per occasion on all 30 days and the R continued to answer the follow-up question AL06DKRE as DK or REF. If the R reported having five or more drinks per occasion on fewer than 30 days, then the value from DR5DAY was retained, and the estimated 30-day frequency variable AL30EST was assigned a bad data value. For example, if an R reported having five or more drinks per occasion on 10 days in the past 30 days but estimated drinking alcohol on 6 to 9 days in that period, it would be impossible for these two answers to be consistent.
R drank on 1 day in the past month and the usual number of drinks was five or more on this 1 day, but frequency of consumption of five or more drinks per occasion is 0 (or is answered as DK/REF).	A value of 1 day was assigned to DR5DAY. If the R reported having five or more drinks on the 1 day when he or she drank, then the R logically had to have had five or more drinks on exactly 1 day in the past month.
R drank on 1 day in the past month and the usual number of drinks was fewer than five, but the frequency of consumption of five or more drinks per occasion is answered as DK/REF.	A code of 80 was assigned to DR5DAY, where 80 = NO OCCASION OF 5+ DRINKS IN PAST 30 DAYS Logically assigned. If the R reported having fewer than five drinks on the 1 day when he or she drank, then the R logically had no occasions of consuming five or more drinks in the past month.
Frequency of consumption of five or more drinks in the past 30 days (DR5DAY) falls within the range of the estimated frequency of consumption of alcohol in the past 30 days (AL30EST).	No further editing was done because the data were not necessarily inconsistent. For example, if the R reported drinking any alcohol on 6 to 9 days in the past 30 days and reported having five or more drinks per occasion on 7 days in that period, these data would be consistent as long as the R drank any alcohol on 7, 8, or 9 days in the past 30 days. In this situation, minimum and maximum possible values were created for use by the statistical imputation team in assigning a final 30-day frequency for alcohol. This procedure narrowed the allowable range of the alcohol 30-day frequency variable ALCDAYS for consistency with the five-drink frequency variable DR5DAY. The final 30-day frequency for alcohol was picked from within that narrowed range. In the above example where the R reported having five or more drinks per occasion on 7 days in the past 30 days and 6 to 9 days in that period, the allowable range for the 30-day frequency would be 7 to 9 days, not 6 to 9, in order for ALCDAYS to be consistent with DR5DAY.

Table B.7 Edits Pertaining to the 30-Day Alcohol Frequency Variables Related to Binge Alcohol Use (continued)

Issue	Edits Implemented
The R drank on more than 1 day in the past month and the usual number of drinks per day was five or more, but the reported frequency of five or more drinks per occasion is 0.	*Both* the usual number of drinks (i.e., NODR30A, from AL07) and DR5DAY (from AL08) were set to bad data. This approach was consistent with cognitive testing results that suggested that Rs sometimes incorrectly answer AL07 by indicating the total number of drinks they had *over the entire 30-day period* instead of the usual number of drinks they had per day. Consequently, these edits did not automatically assume that these Rs were binge alcohol users. If DR5DAY had a value of DK/REF and the usual number of drinks reported was five or more, then DR5DAY retained the corresponding DK/REF value.
The number of days that the R reported having five or more drinks equals the total number of days that the R reported drinking any alcohol (DR5DAY = ALCDAYS), but the reported usual number of drinks is fewer than five (or is answered as DK/REF).	A code of 975 was assigned to NODR30A (i.e., the edited version of AL07) to indicate usual consumption of at least five drinks.

Table B.8 Edits Pertaining to Incidence Variables Based on Consistency Checks

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (e.g., CGCC22 for cigarettes) that the age at first use (AFU) based on the new month and year of first use (MFU and YFU) was correct (e.g., CGCC22 = 4). The computer-assisted interviewing (CAI) program updated the AFU (such as AGE1STCG for cigarettes) with the value of the age calculated from the MFU and YFU (referred to subsequently as the MYR1ST age, such as MYR1STCG, for cigarettes). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first used the drug (e.g., the R was 16, reported first use of the drug at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used the drug). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained for the AFU (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable (e.g., CIGYFU for cigarettes), and a code of 89 was assigned to the MFU variable (e.g., CIGMFU). Consistent with procedures in prior years, therefore, the values for the MFU and YFU were statistically imputed.
The final verification check (e.g., CGCC22) was skipped because the R entered revised data for the MFU and YFU that made them consistent with the AFU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (e.g., CGCC22) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the MFU and YFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 1994, was interviewed in March 2012 (age 17 at the time of the interview), reported first use of a drug at age 17, and initially reported first use in May 2011. First use in May 2011 would have meant that the R was 16 when he or she first used the drug because the R's 17th birthday was not until June 2011. If the R changed the month and year to June 2011, that could be consistent with first use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2010, it would never be possible for the R to have first used at age 17 and also to have first used in June 2010. In this latter situation, the MFU and YFU would be set to bad data.
The R entered a new MFU or YFU that differed from what the R previously reported. The MYR1ST age based on the revised MFU and YFU still mismatched the AFU, but the R indicated in the final verification check that this MYR1ST age was correct.	No editing was done in this situation. The CAI program automatically updated the AFU to be consistent with the updated values reported for the MFU and YFU.

Table B.8 Edits Pertaining to Incidence Variables Based on Consistency Checks (continued)

Issue	Edits Implemented
The consistency check was triggered between the AFU and the MYR1ST age calculated from the MFU and YFU. However, the R answered the first consistency check (e.g., CGCC19 for cigarettes) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU was retained, but the MFU and YFU were set to bad data. Consequently, the MFU and YFU needed to be imputed, but the AFU did not.
The R entered values in the consistency checks for the MFU and YFU that again yielded a nonmissing MYR1ST age based on the MFU and YFU. However, the R failed to resolve the inconsistency between the AFU and the MYR1ST age. The R also reported either in the first verification check that the MYR1ST age was not correct (e.g., CGCC19 = 6 for cigarettes) or reported in the second verification check that the AFU was correct (e.g., CGCC20 = 2 for cigarettes).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set the MFU and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MFU and YFU were assigned the code that corresponded to the answer in the final verification check.
The R entered values in the consistency checks for the MFU and YFU that again yielded a nonmissing MYR1ST age based on the MFU and YFU. However, the R failed to resolve the inconsistency between the AFU and the MYR1ST age. The R also reported in the second verification check that neither the AFU nor original MYR1ST age was correct (e.g., CGCC20 = 3 for cigarettes).	The following edits were implemented for the AFU, MFU, and YFU: • The default edit was to set the AFU, MFU, and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the AFU, MFU, and YFU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the AFU, MFU, and YFU and reported that the MYR1ST age calculated from the MFU and YFU was correct (e.g., CGCC19 = 4). However, the R answered the second consistency check (e.g., CGCC20 for cigarettes) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first used a drug, or whether the MFU and YFU indicated the R's correct age at initiation. Therefore, the AFU, MFU, and YFU all were set to missing values.

Table B.9 Edits Pertaining to Incidence Variables Other than for Parent/Child Relationships

Issue	Edits Implemented
The age at first use (AFU) was greater than the respondent's (R's) reported age; this applied to all drugs except pipe tobacco.	The final age was accepted as the standard, and the inconsistent AFU was set to bad data. Any month and year data associated with the AFU also were set to bad data.
The AFU had a value of "don't know" (DK) or "refused" (REF), including situations where this assignment has been made from the consistency check data.	If the MFU and YFU questions were skipped because the R answered the age at first use as DK or REF to begin with, the DK or REF value from the AFU was propagated onto the skipped MFU and YFU variables. This edit was designed to indicate the reason that the MFU and YFU variables had been skipped. In addition, because the R may have first used the drug within 1 year of his or her current age, the MFU and YFU questions may have been relevant to the R. If the R had answered the MFU and YFU questions but the final age at first use had a value of DK or REF (i.e., due to a consistency check response), the MFU and YFU data were overwritten with the corresponding DK or REF value from the edited age at first use. Retaining the MFU and YFU data in this situation would imply that the R used within 1 year of his or her current age.
The MFU and/or YFU questions were answered because the initial answer to an AFU was within 1 year of the R's current age. However, the final, edited AFU is more than 1 year younger than the R's current age based on the consistency checks between the AFU, YFU, and MFU that were described in Table B.8.	The original answers to the MFU and YFU were overwritten with logically assigned legitimate skip codes (Section 2.4.2). If the R originally had answered the relevant AFU as being more than 1 year younger than his or her current age, the CAI program would have skipped the R past the questions about the MFU and YFU.
The MFU has been skipped because the R answered the YFU as DK or REF.	The DK or REF value from the YFU was propagated onto the skipped MFU. That is, if the R did not know in what year he or she first used the drug, it was assumed that the R would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year.
The R had missing data for the MFU. However, one of the following occurred: • The YFU was the current calendar year, and the R was interviewed in January. • The R first used at his or her current age, the R first used in the current calendar year, and the R's most recent birthday occurred in the interview month. • The R first used at his or her current age, the R first used in the prior calendar year, and the R's most recent birthday occurred in December.	The MFU could be logically inferred, as indicated below. • The MFU was logically inferred to be January of the interview year. That would be the only month in which the R could have initiated use in the current year. • The MFU was logically inferred to be the interview month. If the R first used in the current calendar year and attained his or her current age in the interview month, the R logically could not have initiated use in any month other than the interview month. • The MFU was logically inferred to be December. If the R first used in the prior calendar year and attained his or her current age in December of that year, the R logically could not have initiated use in any month other than December of the prior calendar year.

Table B.10 Edits Pertaining to Parent/Child Incidence Variables

Issue	Edits Implemented
The AFU for a child drug (e.g., LSDAGE for LSD) indicated first use at age 1 or 2, but the AFU for the parent drug (e.g., HALLAGE for any hallucinogen) had a missing value. This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/ OxyContin®, and stimulants/methamphetamine.	Ages of 1 or 2 in the AFU variables for child drugs (e.g., LSDAGE) were set to bad data. The imputation procedures for AFU variables did not allow Rs with AFUs of 1 or 2 to be donors for imputation. Logically, however, an AFU of 1 or 2 for a child drug would preclude the AFU for the parent drug from being imputed to any value other than 1 or 2. For example, if LSDAGE was 1 or 2 and the AFU for any hallucinogen (HALLAGE) had a missing value, the only potential imputation donors for HALLAGE would be those Rs with HALLAGE of 1 or 2. By definition, however, no one could be a donor for imputation. By setting LSDAGE to bad data, the AFUs for both LSD and any hallucinogen would be imputed.
The AFU for a child drug (e.g., LSDAGE) indicated first use at age 3, but the AFU for the parent drug (e.g., HALLAGE) had a missing value. This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/OxyContin®, and stimulants/methamphetamine.	No editing was done when this pattern occurred. Logically, however, the AFU for the parent drug had to have a value of 3 because respondents with AFUs of 1 or 2 (e.g., HALLAGE of 1 or 2) were not eligible to be donors in the imputations. The final imputed AFU (e.g., IRHALAGE for any hallucinogen) was set to a value of 3.
The AFU for a child drug (e.g., crack cocaine) was younger than AFU for the parent (e.g., any cocaine), This applied to first use of cigarettes and age at initiation of daily smoking, and to AFUs for cocaine/crack, hallucinogens/LSD/PCP/ Ecstasy, pain relievers/OxyContin®, and stimulants/ methamphetamine.	This edit required the AFUs to be defined (i.e., not missing) for both the parent and child drugs. The younger AFU from the child drug was assigned to the parent AFU. For example, if the AFU of crack cocaine was younger than the AFU of any cocaine, then the edits assigned the value from the crack AFU to the cocaine AFU.
Respondents (Rs) in the stimulants module were lifetime users only of methamphetamine. This included situations in which the only reported use of a stimulant was "some other stimulant," but only methamphetamine was specified. The computer-assisted interviewing (CAI) program skipped these respondents out of questions related to general stimulant use.	Methamphetamine data were moved over to the corresponding general stimulant variables.
Rs in the pain relievers module were lifetime users only of OxyContin®. This included situations in which the only reported use of a pain reliever was "some other pain reliever," but only OxyContin® was specified. The CAI program skipped these respondents out of general questions related to use of any pain relievers.	OxyContin® data were moved over to the corresponding general variables for any pain relievers.

Table B.10 Edits Pertaining to Parent/Child Incidence Variables (continued)

Issue	Edits Implemented
Rs in the hallucinogens module were lifetime users only of LSD, only of PCP, or only of Ecstasy. This included situations in which the only reported use of a hallucinogen was "some other hallucinogen," but only LSD, only PCP, or only Ecstasy was specified. The CAI logic skipped these respondents out of questions related to the age, month, and year that they first used the specific hallucinogen.	Data on initiation of hallucinogen use were moved over to the corresponding specific hallucinogen variables that had been skipped. For example, if the R had used only LSD, the hallucinogen age (from HALLAGE) was assigned to the LSD AFU (LSDAGE). Similarly, if the R had provided data on the month and year that he or she had first used any hallucinogen, these data were assigned to the LSD month and year variables LSDMFU and LSDYFU. Similar edits were done when Rs had been skipped out of questions related to specific hallucinogens but were logically inferred to be lifetime users only of LSD, only of PCP, or only of Ecstasy, based on the "OTHER, Specify" responses.
Rs in the hallucinogens module were lifetime users of two or more of the following: LSD, PCP, or Ecstasy. No other hallucinogen use was indicated. However, the R answered the AFU question for any hallucinogen (edited variable HALLAGE) as "don't know" or "refused." In addition, the specific hallucinogens that the R had used had legitimate values for his or her AFU variables (i.e., LSDAGE, PCPAGE, or ECSAGE, for LSD, PCP, or Ecstasy, respectively).	Logically, the AFU for any hallucinogen (HALLAGE) had to be the minimum of the ages indicated in LSDAGE, PCPAGE, or ECSAGE. Therefore, HALLAGE was assigned the minimum value from LSDAGE, PCPAGE, or ECSAGE. If this edit was done and month and year data were available for a specific hallucinogen, the month and year variables for any hallucinogen (HALMFU and HALYFU) were edited accordingly. For example, if a missing value in HALLAGE was replaced with the value from LSDAGE, and the month of first use (MFU) and year of first use (YFU) for LSD (LSDMFU and LSDYFU) had valid values, the values from LSDMFU and LSDYFU were assigned to HALMFU and HALYFU, respectively. This edit was done *only* to replace missing data in HALLAGE. No editing was done if HALLAGE had a value that was lower than the AFU values reported in LSDAGE, PCPAGE, or ECSAGE.
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first use of any drug in that category (e.g., any hallucinogen) at their current age but had missing data for the age when they first used specific drugs in that category (e.g., LSD).	The Rs were logically inferred to have first used the specific drug at their current age.
The AFU the parent drug was at the R's current age and the AFU for the child drug was at 1 year below the R's current age. Consequently, the child AFU had been assigned to the parent AFU.	The associated MFU and YFU from the child drug (e.g., crack) were assigned to the parent' drug's MFU and YFU data (e.g., for any cocaine). However, this edit was not done in the hallucinogens module when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy. Consequently, some inconsistencies could remain in the edited hallucinogens data.
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first use for the parent drug category (e.g., any hallucinogen) in the current survey year but had missing data for the year when they first used a child drug (e.g., LSD).	The Rs were logically inferred to have first used the child drug in the current year. If the Rs also reported first use of drugs in the parent category (e.g., any hallucinogen) in the same month that they were interviewed and they had missing MFU data for child drugs within that category, the Rs also were logically inferred to have first used the child drug in the same month that they were interviewed. Otherwise, if the Rs originally had reported a month when they first used a child drug but that month had been assigned a code of bad data due to the "flag and impute" edits, the MFU (e.g., LSDMFU) also was reset to the originally reported value if it matched the MFU for any drug in the parent category (e.g., HALMFU).

Table B.10 Edits Pertaining to Parent/Child Incidence Variables (continued)

Issue	Edits Implemented
The Rs in the cocaine/crack, hallucinogens, pain relievers, or stimulants sections indicated first using any drug in that category (e.g., any hallucinogen) in the year immediately prior to the current one (i.e., 2011) but that point of initiation was within 12 months of the interview date. These Rs also had missing data for the year and month when they first used specific drugs in that category (e.g., LSD).	If the Rs originally had reported year and month data for when they first used a specific drug but these data had been assigned codes for bad data due to the "flag and impute" edits, this information was restored *only* if the original answers matched the MFU and YFU for any drug in that category. For example, if LSDMFU and LSDYFU had been set to bad data but the original answers for the LSD month and year matched the month and year in HALMFU and HALYFU, LSDMFU was equated to HALMFU, and LSDYFU was equated to HALYFU.
The MFU and/or YFU questions for the parent drug were answered because the initial answer to the parent AFU was within 1 year of the R's current age. However, the final, edited AFU for the parent drug was more than 1 year younger than the R's current age (e.g., if the unedited AFU for cocaine was within 1 year of the R's current age, but the AFU of crack was more than 1 year younger than the R's current age).	The original answers to the parent MFU and YFU were overwritten with logically assigned legitimate skip codes (see Section 2.4.2). If the R originally had answered the parent AFU as being more than 1 year younger than his or her current age, the CAI program would have skipped the R past the questions about the parent MFU and YFU.
The AFUs for the parent and child drugs were the same, but the MFU or YFU was earlier for the child drug (e.g., crack).	The earlier month and year data from the child drug (e.g., crack) were assigned to the parent drug's month and year data (e.g., cocaine). However, this edit was not done in the hallucinogens module when the sole hallucinogen use involved use of LSD, PCP, or Ecstasy but the YFU and MFU data for the parent and child drugs were inconsistent. Consequently, some inconsistencies could remain in the edited hallucinogens data.
For cigarettes, the AFU for any cigarette use (CIGTRY) equaled the age at initiation of daily cigarette use (CIGAGE). However, the R reported initiating daily cigarette use in a year or month that was earlier than what the R reported for first use of any cigarette.	The month and year data for daily cigarette use (CIGDLMFU and CIGDLYFU, respectively) were assigned to the month and year variables for any cigarette use (CIGMFU and CIGYFU, respectively).

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module

Issue	Edits Implemented
The respondent (R) reported methamphetamine use in the core stimulants module (i.e., question ST01 or STREF1 was answered as "yes").	The edited recency variables MTHAREC (corresponding to question SD17B) and MTANDLRC (SD18A) were assigned legitimate skip codes. The edited variables MTHAMP (corresponding to question SD17A), MTHANEDL (SD18A), MTHEVCK (from SD17A1 and SD17ALT), MTHNORSN (SD17A2), and MTHNOSP (SD17A2sp) also were assigned legitimate skip codes (see Section 2.4.2).
The R had been routed to the methamphetamine follow-up questions in the special drugs module because the R had not reported methamphetamine use in ST01 or STREF1. However, the R had specified use of methamphetamine as "some other stimulant" in the core stimulants module.	If the lifetime methamphetamine question SD17A had not been answered as "yes," and no other reports of methamphetamine use existed in the special drugs module, then MTHAMP was assigned the code for bad data (see Section 2.4.3).
The R reported in question SD17A that he or she had never used methamphetamine, but the R also specified use of methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The R was logically inferred to be a methamphetamine user. The edited variables MTHAMP and MTHANEDL were assigned a code of 3 (Yes, LOGICALLY ASSIGNED). Because questions SD17B and SD18B pertaining to the most recent use of methamphetamine and most recent use of methamphetamine with a needle, respectively, had been skipped, the edited variables MTHAREC and MTANDLRC were assigned codes of 9 (Used at some point in the lifetime, LOGICALLY ASSIGNED).
After reporting methamphetamine use in SD17A, the R reported in question SD17A1 or SD17ALT that his or her earlier report from the core stimulants module was correct that he or she had never used methamphetamine. However, the R also specified using methamphetamine with a needle.	The "OTHER, Specify" data overruled the R's denial of methamphetamine use from SD17A1 or SD17ALT. MTHAMP retained a code of 1 (i.e., "yes"), and the data were retained from SD17B for MTHAREC. The variable MTHEVCK was assigned the code for bad data. If SD18A did not indicate use of methamphetamine with a needle, MTHANEDL was assigned a code of 3 and MTANDLRC was assigned a code of 9, as described previously.
After reporting methamphetamine use in SD17A, the R reported in question SD17A1 or SD17ALT that his or her earlier report from the core stimulants module was correct that he or she had never used methamphetamine. The R did not specify any use of methamphetamine with a needle.	The R was logically inferred *not* to be a methamphetamine user. MTHAMP was assigned a code of 4 (No, LOGICALLY ASSIGNED), and codes of 81 (Never used methamphetamine, LOGICALLY ASSIGNED) were assigned to MTHAREC, MTHANEDL, and MTANDLRC. The answer that confirmed that the R had never used methamphetamine was retained in MTHEVCK to document the source of the logical inference of nonuse of methamphetamine. MTHNORSN and MTHNOSP were assigned codes of 99 (LEGITIMATE SKIP).
The R did not know or refused to report in SD17A1 or SD17ALT which answer was correct: the earlier report of nonuse from the core stimulants module or the report of methamphetamine use in the special drugs module.	MTHAMP through MTANDLRC were set to bad data (i.e., including MTHAREC). Data from the core stimulants module indicating lifetime nonuse of methamphetamine were used to determine the core-plus-noncore (CPN) methamphetamine recency.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R did not indicate methamphetamine use in the core stimulants module and also reported in SD17A that he or she never used methamphetamine. No other reports of methamphetamine use existed elsewhere in the special drugs module.	MTHAMP retained a code of 2 (i.e., "no"). Codes of 91 (NEVER USED METHAMPHETAMINE) were assigned to MTHAREC, MTHANEDL, and MTANDLRC. Codes of 99 were assigned to MTHEVCK, MTHNORSN, and MTHNOSP.
Follow-up questions were skipped because the R refused to indicate in SD17A whether he or she ever used methamphetamine.	Codes of 97 (i.e., refused) were assigned to the remaining edited methamphetamine variables, and they were treated as having missing data. Thus, a refusal to report methamphetamine use in SD17A was propagated onto MTHAREC, MTHANEDL, and MTANDLRC.
The R reported in SD17A1 or SD17ALT that his or her report of methamphetamine use in question SD17A was correct. The R reported in SD17A2 that there was "some other reason" why he or she had not previously reported methamphetamine use in the core stimulants module. However, the written response in SD17A2SP indicated either that the R did not recognize methamphetamine as a prescription drug, or that the R made a mistake in answering the previous core question about methamphetamine.	MTHNORSN (corresponding to question SD17A2) was edited as follows: • Assigned a code of 11 (Did not consider methamphetamine to be a prescription drug LOGICALLY ASSIGNED) if the R gave an answer in question SD17A2SP indicating that he or she did not recognize methamphetamine in the context of questions about prescription drugs. MTHNOSP (corresponding to SD17A2SP) retained a code of 1 ("Did not think of methamphetamine as a prescription drug") to indicate the source of the logical inference. • Assigned a code of 12 (Made a mistake in answering the previous question LOGICALLY ASSIGNED) if the R gave an answer in SD17A2SP indicating that he or she made a mistake in answering the methamphetamine question in the core stimulants module. MTHNOSP retained a code of 2 (i.e., Made a mistake). Values of 11 and 12 in MTHNORSN were used in editing MTHREC06 (see Section 6.2.6.2).
MTHAMP indicated that the R was a lifetime user of methamphetamine, and there was nothing in the follow-up questions to indicate that this answer was incorrect. However, the R did not know or refused to report when he or she last used methamphetamine.	The methamphetamine recency variable (MTHAREC) was generally assigned a code of 9 (i.e., used at some point in the lifetime). Exceptions were as follows: • If the R reported last using methamphetamine with a needle in the past 30 days, then MTHAREC was assigned a code of 11 (Used in the past 30 days, LOGICALLY ASSIGNED). • Otherwise, if the R reported last using methamphetamine with a needle more than 30 days ago but within the past 12 months, then MTHAREC was assigned a code of 8 (Used at some point in the past 12 months, LOGICALLY ASSIGNED). A code of 8 was assigned because the R could have used methamphetamine in the past 30 days.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R indicated more recent use of methamphetamine with a needle in MTANDLRC than the R indicated for most recent use of any methamphetamine in MTHAREC.	MTHAREC was edited to reflect the indication of more recent use from MTADNLRC. • If MTANDLRC indicated that the R last used methamphetamine with a needle in the past 30 days, but MTHAREC did not indicate use in the past 30 days, MTHAREC was assigned a code of 11 (Used in the past 30 days LOGICALLY ASSIGNED). • If MTANDLRC indicated that the R last used methamphetamine with a needle more than 30 days ago but within the last 12 months, and MTHAREC indicated that the R last used methamphetamine more than 12 months ago, MTHAREC was assigned a code of 12 (Used more than 30 days ago but within the past 12 months LOGICALLY ASSIGNED). • If MTANDLRC indicated that the R last used methamphetamine with a needle more than 30 days ago but within the last 12 months, and MTHAREC did not indicate use in the past year, MTHAREC was assigned a code of 8 (Used at some point in the past 12 months LOGICALLY ASSIGNED.
MTHAMP (corresponding to question SD17A) indicated that the R was a lifetime user of methamphetamine. There were no indications elsewhere in the special drugs module of the R reporting use of methamphetamine as "some other drug" that he or she used with a needle. The corresponding needle recency variable MTANDLRC was skipped because of the following:	
• The R never used methamphetamine with a needle (from MTHANEDL).	Codes of 93 were assigned to the corresponding methamphetamine needle recency variable MTANDLRC to indicate that the R used methamphetamine but never with a needle.
• The R refused to indicate or did not know in MTHANEDL whether he or she had ever used methamphetamine with a needle.	A code of 97 was assigned to MTANDLRC (i.e., the refusal was propagated). Data from MTANDLRC were not relevant to creating the CPN methamphetamine recency.
The R was a lifetime user of methamphetamine or other stimulants in the core stimulants module and reported using methamphetamine or other stimulants with a needle at least once in his or her lifetime. However, the R did not know or refused to indicate when he or she last used the drug with a needle.	The edits depended on the most recent use reported in the corresponding core recency variable: • In general, the edited needle recency variable (e.g., MTNDLREC for methamphetamine, based on question SD10B) was assigned a code of 9 (i.e., used at some point in the lifetime). • However, if the core recency (e.g., METHREC) indicated that the R last used the drug more than 12 months ago, then the edited needle recency variable pertaining to that drug was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R was a lifetime methamphetamine user in the core stimulants module. The R reported in question SD10A that he or she had never used a needle to inject methamphetamine, but the R also specified use of methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The lifetime methamphetamine needle use variable MTHNEEDL (corresponding to SD10A) was assigned a code of 3 (Yes LOGICALLY ASSIGNED). The methamphetamine needle recency variable MTNDLREC generally was assigned a code of 9 (i.e., used at some point in the lifetime). However, if the R reported last use of any methamphetamine more than 12 months ago (i.e., METHREC = 3), then MTNDLREC was assigned a code of 13 (Used a needle to inject methamphetamine more than 12 months ago, LOGICALLY ASSIGNED).
The R was a lifetime methamphetamine user in the core stimulants module. Questions SD10C and SD11 pertaining to use of other stimulants with a needle had been skipped because methamphetamine was the only stimulant that the R had reported ever using.	If the lifetime methamphetamine variable METHDES was coded as 1 (i.e., "yes") and all other lifetime stimulant variables had values of 2 (i.e., "no"), the edited needle use variables for lifetime use of other stimulants with a needle (OSTNEEDL, corresponding to SD10C) and most recent use of other stimulants with a needle (OSTNLREC, corresponding to SD11) were assigned a code of 99 (legitimate skip). This edit was not done if SD10C and SD11 were skipped when METHDES had the only affirmative response, but at least one of the other lifetime stimulant variables had a value of "don't know" or "refused."
The R was a lifetime methamphetamine user in the core stimulants module. In addition, questions SD10C and SD11 pertaining to use of other stimulants with a needle had been answered because the R reported use of "some other stimulant" in the stimulants module. Based on "OTHER, Specify" data in the stimulants module, however, methamphetamine was the only stimulant that the R had ever used.	Data were retained in the needle recency variable for other stimulants (OSTNLREC) if the R reported more recent use of "other" stimulants with a needle than what the R reported for the methamphetamine needle recency (MTNDLREC); that is, no editing was done. Otherwise, nonblank values in OSTNLREC were replaced with codes of 89 (Legitimate skip LOGICALLY ASSIGNED) (see Section 2.4.2).
The R was a lifetime nonmedical user of stimulants. The R reported in question SD10C that he or she had never used a needle to inject stimulants/other stimulants. However, the R also specified use of stimulants other than methamphetamine with a needle somewhere in the "OTHER, Specify" items SD05A through SD05E, pertaining to use of other drugs with a needle.	The stimulant needle recency variable OSTNLREC generally was assigned a code of 9 (i.e., used at some point in the lifetime). However, if the R reported last nonmedical use of any stimulant more than 12 months ago (i.e., STIMREC = 3), then OSTNLREC was assigned a code of 13 (Used a needle to inject stimulants more than 12 months ago, LOGICALLY ASSIGNED).
The R was asked questions about use of methamphetamine and other stimulants with a needle because the R reported lifetime use of methamphetamine and "some other stimulant" in the stimulants module (and no other stimulant). However, only methamphetamine was specified as the "other" stimulant.	The R was treated as being a lifetime user only of methamphetamine. Therefore, any data in the other stimulant needle use variables OSTNEEDL and OSTNLREC were replaced with codes of 89 (LEGITIMATE SKIP Logically assigned). This edit indicated that the R logically should have skipped the questions pertaining to OSTNEEDL and OSTNLREC.

Table B.11 Edits Pertaining to Variables for Methamphetamine Prevalence, Stimulant Prevalence, and Related Follow-Up Variables in the Special Drugs Module (continued)

Issue	Edits Implemented
The R had been logically inferred to be a nonuser of prescription-type stimulants because the only drugs that the R reported using in the stimulants module were over-the-counter (OTC) drugs. In addition, the R did not specify use of methamphetamine or other stimulants with a needle as "some other drug" that the R used with a needle.	Any data in the methamphetamine and other stimulant needle variables MTHNEEDL, MTNDLREC, OSTNEEDL, and OSTNLREC were overwritten with a code of 81 (i.e., NEVER USED METHAMPHETAMINE/STIMULANTS Logically assigned) for consistency with the inference that the R was a lifetime nonuser of prescription-type stimulants. This edit did not apply to the methamphetamine variables starting with question SD17A that respondents were asked if they reported in the core stimulants module that they never used methamphetamine.
The R reported in the core drug modules that he or she never used methamphetamine (METHREC = 81 or 91) or that he or she had never used stimulants (STIMREC = 81 or 91). The R also had not reported use of methamphetamine in question SD17A. However, the R specified use of methamphetamine or other stimulants as "some other drug" that he or she had ever injected.	No editing was done to the core drug data, including the core recency variables. However, the edits that were described previously for setting the relevant needle recency variable to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED) were relevant to creating the CPN recency variables for stimulants and methamphetamine.
The R was logically inferred to be a lifetime user of methamphetamine with a needle in the MTHNEEDL variable that existed prior to 2005 (i.e., MTHNEEDL = 3) based on the R's "OTHER Specify" data in the variables OTDGNDLA through OTDGNDLE.	If the lifetime other stimulant needle use variable OSTNEEDL had missing data regarding use of a needle to inject other stimulants, whatever unedited data existed in questions SD10C and SD11 were reassigned to the corresponding edited variables OSTNEEDL and OSTNLREC, respectively.
The R was logically inferred to be a lifetime user of methamphetamine (METHDES = 3) based on "OTHER, Specify" data in the stimulants module, and the methamphetamine needle variable MTHNEEDL had missing data. Further, the variable MTHANEDL did not indicate use of methamphetamine with a needle.	If OSTNEEDL = 2 and MTHANEDL did not equal 1 (i.e., "yes"), the R was logically inferred not to have used a needle to inject methamphetamine (i.e., MTHNEEDL = 4, where 4 = No LOGICALLY ASSIGNED). If the methamphetamine needle recency variable MTNDLREC had been skipped, it was assigned a code of 93 (USED METHAMPHETAMINE BUT NEVER WITH A NEEDLE). When the R had not indicated lifetime methamphetamine use in questions ST01 or STREF1 in the core stimulants module, question SD10C (corresponding to OSTNEEDL) asked whether the R had used a needle to inject *any* stimulant. Therefore, it could be logically inferred that the R had never used a needle to inject methamphetamine. Consequently, MTNLDREC would not indicate more recent use of methamphetamine with a needle than was indicated by METHREC.

Table B.12 Miscellaneous Core Edits Pertaining to Tobacco Variables

Issue	Edits Implemented
The R reported smoking cigarettes on all 30 days in the past 30 days, but the R reported initiating daily cigarette use within 30 days of the interview date. The latter could occur in one of the following ways: (a) the reported month and year for when the R started smoking cigarettes daily was within 30 days of the interview date; or (b) the R reported starting to smoke cigarettes every day at his or her current age (i.e., CIGAGE = AGE), but the R had been at this age for fewer than 30 days.	The following edits were implemented: • If only the month and year indicated that the R had been smoking daily for fewer than 30 days (i.e., but the R had been at the age given in CIGAGE for at least 30 days), the month and year (CIGDLMFU and CIGDLYFU) were set to bad data. • If the R had been at the age reported in CIGAGE for fewer than 30 days and reported starting to smoke daily in the same month that he or she was interviewed, CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data. • If the R had been at the age reported in CIGAGE for fewer than 30 days and CIGDLMFU and CIGDLYFU had codes of DK or REF, CIGAGE, CIGDLMFU, and CIGDLYFU all were set to bad data.
The R reported smoking cigarettes in the past 30 days. The R did not report smoking on all 30 days in the past 30 days but reported initiating daily cigarette use within 30 days of the interview date.	No editing was done when this pattern occurred, even though question CG15 asked respondents if they ever smoked cigarettes every day for at least 30 days. Consequently, data were preserved that indicated that the R recently initiated daily cigarette use, even if the R could not have smoked daily for at least 30 days.
The CAI program had skipped lifetime cigarette questions pertaining to periods where the R had smoked cigarettes daily for at least 30 days (CG15) or whether the R had smoked 100 or more cigarettes (CG16A). Answers to prior questions already indicated that the R had engaged in these behaviors. For example, if the R had reported smoking every day in the past 30 days, the R had to have had a point in his or her life when he or she smoked cigarettes every day for at least a month. Similarly, it could be inferred from answers to questions about the number of days that Rs smoked in the past 30 days and the number of cigarettes they usually smoked per day that they had smoked at least 100 cigarettes in their lifetime.	It was logically inferred in CIGDLYMO (the edited version of CG15) or in CIG100LF (the edited version of CG16A) that the R had engaged in the behavior of interest (i.e., ever smoked cigarettes daily for at least 30 days or smoked 100 or more cigarettes in the R's lifetime). A code of 5 was assigned. This code had the following meaning: 5 = Yes LOGICALLY ASSIGNED (from skip pattern) This code was used instead of a legitimate skip code to signify that the CAI logic had skipped the R out of the question because the R had already provided data to indicate that the question should have been answered affirmatively.
The R was routed to the lifetime 100-cigarette question (CG16A) because the R estimated the number of days that he or she smoked cigarettes in the past 30 days. The R did not answer CG16A as "yes," but the minimum number of days that the R could have smoked in the past 30 days and the minimum usual number of cigarettes smoked per day suggest that the R smoked 100 or more cigarettes in his or her lifetime.	It was logically inferred in CIG100LF (i.e., the edited version of CG16A) that the R had smoked 100 or more cigarettes in his or her lifetime. For example, if an R reported smoking 10 to 19 days and reported smoking 16 to 25 cigarettes per day, then the R would have smoked at least 100 cigarettes in the past 30 days, even if he or she smoked on only 10 days and smoked only 16 cigarettes per day.

B.3 Detailed Documentation of Edits for Variables in Noncore Self-Administered Modules

This section presents tables that detail the edits that were applied to variables in the noncore self-administered modules. Because the content of noncore self-administered modules differs considerably from module to module, many of the edits are module-specific.

- Tables B.13 and B.14 present edits pertaining to variables in the special drugs module (except for issues that were presented in Table B.12 for editing of stimulant and methamphetamine variables from this module), corresponding to the discussion of procedures in Section 7.4.1.
- Tables B.15 and B.16 present edits pertaining to variables in the blunts module, corresponding to the discussion of procedures in Section 7.4.3.
- Table B.17 presents edits pertaining to variables in the special topics module, corresponding to the discussion of procedures in Section 7.4.5.
- Tables B.18 to B.21 present edits pertaining to variables in the prior substance use module, corresponding to the discussion of procedures in Section 7.4.7.
- Tables B.22 to B.24 present edits pertaining to variables in the substance treatment module, corresponding to the discussion of procedures in Section 7.4.8.
- Table B.25 presents edits pertaining to editing of variables in the adult mental health service utilization module, corresponding to the discussion of procedures in Section 7.4.10.
- Finally in this section, Tables B.26 and B.27 present edits pertaining to variables in the consumption of alcohol module, corresponding to the discussion of procedures in Section 7.4.17.

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module

Response Pattern	Edit
Variables were skipped because the respondent (R) never used the drug of interest, and there were no other indications elsewhere in the special drugs module that the respondent ever used this drug.	Codes of 91 were assigned to the edited variables. For example, if the R never used heroin, the edited variables HERSMOKE, HRSMKREC, HERSNIFF, HRSNFREC, HERNEEDL, and HRNDLREC were assigned codes of 91.
Variables were skipped because the R refused to indicate in the corresponding core module whether he or she ever used the drug of interest, and there were no other indications elsewhere in the special drugs module that the respondent ever used this drug.	Codes of 97 (i.e., refused) were assigned to the edited variables. Thus, for example, a refusal from the heroin recency-of-use variable in the core was propagated onto the heroin variables in the special drugs module.
Variables were skipped because the R did not know in the core module whether he or she ever used the drug of interest, and there were no other indications elsewhere in the special drugs module that the R ever used this drug.	The skipped special drugs variables pertaining to this drug retained a value of 98 (i.e., blank).
For the methamphetamine variables that existed prior to 2005, other stimulants, and cocaine, there were no indications elsewhere in the special drugs module of the R reporting use of this drug as "some other drug" that he or she used with a needle. The R was a lifetime user of the drug of interest, but the corresponding needle recency variable was skipped because of the following:	
• The R never used the drug with a needle.	Codes of 93 were assigned to the corresponding needle recency variable (e.g., CONDLREC) to indicate that the R used the drug but never with a needle.
• The R refused to indicate whether he or she had ever used the drug with a needle.	A code of 97 was assigned to the edited needle recency variable (i.e., the refusal was propagated).
• The R did not know whether he or she had ever used that drug with a needle.	The edited needle recency variable retained a code of 98 (i.e., blank).

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module (continued)

Response Pattern	Edit
The R was a lifetime user of heroin, but relevant recency variables for smoking heroin (HRSMKREC), sniffing heroin (HRSNFREC), or using it with a needle (HRNDLREC) were skipped because of the following:	
• The R never used heroin via the route of interest.	Codes of 93 were assigned to the relevant heroin recency variable(s) (e.g., HRSMKREC for smoking heroin) to indicate that the R used heroin but not in that particular way.
• The R refused to indicate whether he or she had ever used heroin via the route of interest.	A code of 97 was assigned to the relevant heroin recency variable(s) (i.e., the refusal was propagated).
• The R did not know whether he or she had ever used heroin via the route of interest.	The edited heroin recency variable(s) retained a code of 98 (i.e., blank).
(In the case of heroin use with a needle, there were no other indications elsewhere in the special drugs module of heroin use with a needle.)	The special situation in which Rs reported lifetime use of heroin in the core but reported that they never smoked, sniffed, or used it with a needle is discussed in the text.
General needle use variables were skipped because the R reported never using heroin, methamphetamine, other stimulants, cocaine, or any other drug (question SD05; edited variable OTDGNEDL) with a needle. For methamphetamine, this included reports in the methamphetamine question SD18a (added in 2005) and in the methamphetamine question that existed prior to 2005 that the R had never used methamphetamine with a needle.	Codes of 99 (i.e., legitimate skip) were assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET).
General needle use variables were skipped because question SD05 was answered as "no" (OTDGNEDL = 2); there were no affirmative reports of heroin, methamphetamine, other stimulants, or cocaine with a needle, but one or more of the lifetime needle use variables for these drugs was answered as "don't know" or "refused."	When there was no affirmative report of use of heroin, methamphetamine, other stimulants, or cocaine with a needle, question SD05 was worded as follows: "Have you ever, even once, used a needle to inject *any drug* that was not prescribed for you ..." (wording not italicized in the interview). Therefore, codes of 99 (i.e., legitimate skip) were assigned to all of the general needle use variables that had been skipped (GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET) because it could be inferred that the response of "no" in question SD05 pertained to use of any drug with a needle. However, no editing was done to any responses of "don't know" or "refused" in the lifetime needle use variables pertaining to heroin (HERNEEDL), methamphetamine (MTHNEEDL), other stimulants (OSTNEEDL), or cocaine (COCNEEDL).

Table B.13 Specific Skip Logic Edits for Heroin Use and Needle Use Variables in the Special Drugs Module (continued)

Response Pattern	Edit
General needle use variables were skipped because question SD05 was refused (OTDGNEDL = 97), and there were no affirmative reports of heroin, methamphetamine, other stimulants, or cocaine with a needle.	The refusal from OTDGNEDL was propagated to the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use

Issue	Edits Implemented
The respondent (R) was a lifetime user of heroin and reported smoking, sniffing, or using heroin with a needle at least once in his or her lifetime, but did not know or refused to indicate when he or she last smoked, sniffed, or injected heroin.	The edits depended on the most recent use of heroin reported in the corresponding core heroin recency variable: • In general, the edited heroin recency variables in the special drugs module (HRSMKREC, HRSNFREC, HRNDLREC) were assigned a code of 9 (i.e., used at some point in the lifetime). • However, if the core heroin recency indicated that the R last used heroin more than 12 months ago and there was no other indication in the special drugs module that the R had used heroin in the past 12 months (see below), then the edited variables pertaining to smoking, sniffing, or injection of heroin were assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED). This edit did not apply if the R answered "did not know" or refused to report when he or she last used heroin in a particular way (e.g., smoking it) but reported last using it a different way in the past 12 months (e.g., with a needle).
The R was a lifetime user of cocaine and reported using it with a needle at least once in his or her lifetime. However, the R did not know or refused to indicate when he or she last used cocaine with a needle.	The edits depended on the most recent use reported in the corresponding core recency variable COCREC: • In general, the edited needle recency variable CONDLREC for cocaine was assigned a code of 9 (i.e., used at some point in the lifetime). • However, if COCREC indicated that the R last used cocaine more than 12 months ago, then CONDLREC was assigned a code of 13 (i.e., More than 12 months ago LOGICALLY ASSIGNED).
The R reported in the core drug modules that he or she never used heroin (HERREC = 91) or had never used cocaine (COCREC = 91). However, the R specified use of one or more of these drugs as "some other drug" that he or she had ever injected.	No editing was done to the core drug data. However, the R was logically inferred in the special drugs data to be a lifetime user of that drug with a needle, even though the core drug data indicated that the R never used that drug. The corresponding needle recency variable was set to a value of 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). For example, if the R reported in the heroin module that he or she never used heroin but specified injection of heroin as "some other drug," the lifetime heroin needle use variable HERNEEDL was set to 3 (Yes LOGICALLY ASSIGNED), and the heroin needle recency HRNDLREC was set to 9. Similar edits were done for the needle use variables pertaining to methamphetamine, other stimulants, and cocaine.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use (continued)

Issue	Edits Implemented
The R specified lifetime use of more than five drugs with a needle.	Priority was given to retaining as many unique mentions as possible for other drugs that the R used with a needle. Thus, multiple mentions of the same drug would be counted only once. Priority also would be given to retaining mentions of drugs that were covered in the special drugs module that the R had not previously reported using with a needle (e.g., if the question corresponding to MTHNEEDL had been answered as "no" but methamphetamine had been specified as "some other drug" that the R used with a needle). Conversely, retention of "OTHER, Specify" codes corresponding to drugs that the R had already reported using with a needle were given lower priority. If there were still more than five mentions of unique drugs after the above steps, priority was given to retaining the most serious drugs according to the Drug Enforcement Administration (DEA) drug schedule (e.g., first priority to retaining mention of Schedule I drugs that have no approved medical use in the United States, second priority to retaining Schedule II drugs). Finally, after the drugs had been ranked according to their severity based on the DEA drug schedule, if mention of more than five drugs still remained, the codes were retained in the order they appeared in the data.
The R reported using a needle to inject a drug for nonmedical reasons (SD05 = 1), but the R previously reported never using marijuana, cocaine, heroin, hallucinogens, inhalants, prescription pain relievers, prescription tranquilizers, prescription stimulants, or prescription sedatives.	No editing was done if the R • specified needle use involving a drug that could be abused or had psychoactive properties (e.g., steroids, one or more categories of drugs covered in the core modules that were not covered elsewhere in special drugs, such as prescription pain relievers), or • reported one or more "risky" needle use behaviors (reusing a needle, needle sharing, or cleaning a needle with bleach). The R was inferred not to be a lifetime nonmedical needle user (OTDGNEDL = 4) if the R • specified use of a drug that was typically not abused and did not have psychoactive properties (e.g., if injection of antibiotics was specified), and • reported never reusing a needle, sharing a needle (before or after someone had used it), or cleaning a needle with bleach (i.e., "risky" needle use behaviors). When OTDGNEDL = 4, any data in the general needle use variables GNNDREUS, GNNDLSH1, GNNDCLEN, GNNDLSH2, and GNNDGET were replaced with a code of 89 (LEGITIMATE SKIP Logically assigned).
The general needle use items corresponding to questions SD12 through SD16 had been answered. However, the only indications of lifetime needle use had been set to bad data (e.g., if MTHANEDL had been set to bad data; see Table B.11).	The edited general needle use variables GNNDREUS through GNNDGET were set to bad data.

Table B.14 Edits in the Special Drugs Module (Other than Skip Patterns) Pertaining to Heroin Use and Needle Use (continued)

Issue	Edits Implemented
The R reported getting his or her last needle "some other way" and specified a meaningful response for how he or she last got the needle.	The final, edited variable pertaining to how the R got his or her last needle (GNNDGET) was a composite of the response categories that were offered to the R (i.e., bought the needle from a pharmacy, got the needle from a needle exchange, bought the needle on the street, got the needle in a shooting gallery, got the needle some other way). This was done because the computer-assisted interviewing (CAI) logic did not allow Rs to specify an "other" way that they got the needle if they reported getting the needle in one of the first four ways. If Rs reported getting the needle "some other way" and specified a meaningful way they got the needle, then the "OTHER, Specify" response was assigned to GNNDGET.
The R reported getting his or her last needle "some other way" and did not know what that other way was, refused to specify what that other way was, or gave a response that was coded as bad data (e.g., a nonsensical response).	The final, edited variable pertaining to how the R got his or her last needle (GNNDGET) retained a nonspecific code of "some other way." Stated another way, the response of "some other way" was given precedence over the missing value in the "OTHER, Specify" response. The edit was done in this manner to provide a nonmissing value for analysts to use.
The R answered "don't know" or "refused" at the outset, when asked how he or she got the last needle that he or she used.	The response of "don't know" or "refused" was retained in the final, edited variable (GNNDGET).

Table B.15 Edits in the Blunts Module Pertaining to Consistency Checks for Incidence

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (BLCC06) that the age at first use (AFU) based on the new month and year of first use (MFU and YFU) was correct (i.e., BLCC06 = 4). The computer-assisted interviewing (CAI) program updated the AFU (AGE1STBL) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBL). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first used blunts (e.g., the R was 16, reported first use of blunts at age 15, but then confirmed an MFU and YFU that meant the R was 14 when the R first used blunts). Had the R initially reported this AFU, the R would not have been routed to the MFU and YFU questions.	The updated value was retained in the edited AFU variable BLNTAGE (e.g., first use at age 14 for a 16-year-old R in this example). Based on this updated AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the YFU variable BLNTYFU, and a code of 89 was assigned to the MFU variable BLNTMFU.
The final verification check (BLCC06) was skipped because the R entered revised data for the MFU and YFU that made BLNTMFU and BLNTYFU consistent with the AFU in BLNTAGE.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (BLCC06) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, BLNTMFU and BLNTYFU were set to bad data if they could never be consistent with the AFU. Suppose, for example, that a hypothetical R was born in June 1994, was interviewed in March 2012 (age 17 at the time of the interview), reported first use of blunts at age 17, and initially reported first use in May 2011. First use in May 2011 would have meant that the R was 16 when he or she first used blunts because the R's 17th birthday was not until June 2011. If the R changed the month and year to June 2011, that could be consistent with first use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2010, it would never be possible for the R to have first used at age 17 and also to have first used in June 2010. In this latter situation, BLNTMFU and BLNTYFU would be set to bad data.
The R entered a new MFU or YFU that differed from what the R previously reported. The age based on the revised MFU and YFU (updated in MYR1STBL) still mismatched the AFU, but the R indicated in the final verification check that the new value from MYR1STBL was correct.	No editing was done in this situation. The CAI program automatically updated AGE1STBL to be consistent with the updated values reported for the MFU and YFU.

Table B.15 Edits in the Blunts Module Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
The consistency check was triggered between AGE1STBL and MYR1STBL. However, the R answered the first consistency check (BLCC03, regarding whether the value in MYR1STBL was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU from AGE1STBL was retained in BLNTAGE, but BLNTMFU and BLNTYFU were set to bad data.
The consistency check was triggered between AGE1STBL and MYR1STBL. However, the R entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to BLNTAGE, but BLNTYFU and BLNTMFU were set to bad data.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, the R failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. The R also reported either in the first verification check that MYR1STBL was not correct (i.e., BLCC03 = 6) or reported in the second verification check that the AFU from AGE1STBL was correct (i.e., BLCC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set BLNTMFU and BLNTYFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," BLNTMFU and BLNTYFU were assigned the code that corresponded to the answer in the final verification check.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for MYR1STBL. However, the R failed to resolve the inconsistency between AGE1STBL and the updated value in MYR1STBL. The R also reported in the second verification check that neither answer was correct for what was originally captured in AGE1STBL and MYR1STBL (i.e., BLCC04 = 3).	The following edits were implemented for BLNTAGE, BLNTMFU, and BLNTYFU: • The default edit was to set BLNTAGE, BLNTMFU, and BLNTYFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," BLNTAGE, BLNTMFU, and BLNTYFU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the AFU, MFU, and YFU and reported that the age from MYR1STBL was correct (i.e., BLCC03 = 4). However, the R answered the second consistency check (BLCC04) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to BLNTAGE, BLNTMFU, and BLNTYFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first used blunts, or whether the MFU and YFU indicated the R's correct age at initiation. Therefore, BLNTAGE, BLNTMFU, and BLNTYFU all were set to missing values.

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence

Issue	Edits Implemented
The respondent (R) reported lifetime use of blunts but did not know or refused to report in BL02 when he or she last used them.	The blunt recency variable BLNTREC (corresponding to question BL02) was set to 9 (Used at some point in the lifetime LOGICALLY ASSIGNED). The 30-day frequency BLNT30DY was left as blank.
The R reported using blunts (BLNTEVER = 1, corresponding to question BL01), but the R previously reported never using marijuana (MJEVER = 2). The consistency check BLNT30C1 (corresponding to question BL03) was triggered because the R reported using blunts in the past 30 days in question BL02 and the R never used marijuana. BLNT30C1 was not triggered for less recent reports of blunt use.	• If BLNT30C1 = 1 (i.e., R has verified that he or she has never used marijuana): – BLNTEVER was set to 4 (No LOGICALLY ASSIGNED). – BLNTREC was set to 81 (NEVER USED BLUNTS Logically assigned). – BLNT30DY was to 91 (if blank) or 81 (if a value needed to be overwritten). – Incidence variables (BLNTAGE, BLNTYFU, BLNTMFU) were assigned codes of 981, 9981, and 81, respectively. – Other assignment of legitimate skip codes applies, as described in the main text. • If BLNT30C1 = 2 (i.e., R has verified past month use of blunts), no editing was done to BLNTEVER or BLNTREC. Also, these noncore data in blunts were not used to edit the core MJEVER variable (see Section 2.4.1).
The R reported using blunts, but the R previously reported never using marijuana. BLNT30C1 was not triggered because the R reported use of blunts that was less recent than the past 30 days.	• BLNTEVER was set to a value of 11 (i.e., "bump" by 10) to signify that the R reported lifetime use of cigars with marijuana in them but reported never using marijuana in the core, and the R was not asked to verify which answer was correct. • No editing was done to the recency values in BLNTREC.
The R reported using blunts in the past 30 days, and the R is a marijuana user. However, the R reported last using marijuana more than 30 days ago but within the past 12 months or more than 12 months ago. The consistency check BLNT30C2 (corresponding to question BL04) was triggered for these cases.	• If BLNT30C2 = 1 (i.e., R has verified that the core marijuana recency is correct): – BLNTREC was set to 19 (More than 30 days ago LOGICALLY ASSIGNED). – BLNT30DY was set to 93 if blank and to 83 (DID NOT USE BLUNTS IN THE PAST 30 DAYS Logically assigned) if the 30-day frequency data needed to be overwritten. – Note, however, that no editing was done to BLNTREC when BLNT30C2 = 1, but the core marijuana recency MJREC had been set to a value of 8 or 9 (Used at some point in the past 12 months or Used at some point in the lifetime, respectively), based on edits in the marijuana module (see Section 6.2.2.3 in Chapter 6). The rationale for doing no editing to BLNTREC was that values of 8 or 9 in MJREC indicated potential use of marijuana in the past month. • If BLNT30C2 = 2 (i.e., R has verified past month use of blunts), no editing was done to BLNTEVER or BLNTREC. Also, these noncore data in blunts were not used to edit the core marijuana recency MJREC (see Section 2.4.1).

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
The R reported last using blunts more than 30 days ago but within the past 12 months (BLNTREC = 2), and the R reported last using marijuana more than 12 months ago (MJREC = 3).	BLNT30C2 (BL04) was not triggered in this situation because the R did not report use of blunts in the past 30 days. • BLNTREC was set to a value of 12 (i.e., "bump" by 10), to signify that the R reported last using blunts more than 30 days ago but within the past 12 months but reported last using marijuana more than 12 months ago, and the R was not asked to verify which answer was correct. • The editing procedures continued to set the 30-day frequency BLNT30DY to 93 because the R was not a past month user of blunts.
The R reported last using blunts more than 30 days ago but within the past 12 months (BLNTREC = 2), but the marijuana recency MJREC was assigned codes of 8 or 9 (Used at some point in the past 12 months or lifetime, respectively).	No editing was done to the blunts variables when MJREC had values of 8 or 9 because BLNTREC = 2 was not necessarily inconsistent with when the R last used marijuana.
Question BL07 (edited variable BLNTNOMJ) was skipped. The R reported in the tobacco module that he or she never smoked a cigar and reported in the blunts module that he or she never smoked a blunt.	BLNTNOMJ was set to a value of 91 (NEVER USED CIGARS/BLUNTS). This included situations in which BLNTEVER had been set to 4 (No LOGICALLY ASSIGNED) and BLNTREC to 81 (NEVER USED BLUNTS Logically assigned) because the R confirmed in BLNT30C2 that he or she never used marijuana.
BLNTNOMJ was skipped, and the R reported lifetime use of blunts but definitely did not smoke blunts in the past 30 days. The R also reported in the tobacco module that he or she did not smoke a cigar in the past 30 days (including situations in which the R reported never smoking a cigar).	BLNTNOMJ was set to value of 93 (DID NOT USE CIGARS/BLUNTS IN THE PAST 30 DAYS). This edit also was implemented if the R reported use of blunts less recently than the past 30 days but reported in the core tobacco module that he or she never smoked a cigar. Even though the cigar and blunts data were inconsistent, BLNTNOMJ still was not applicable.
BLNTNOMJ was skipped, and the R reported lifetime use of blunts but definitely did not smoke blunts in the past 30 days. Unlike the situation above, the R reported in the tobacco module that he or she smoked cigars in the past 30 days.	BLNTNOMJ was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern). For instance, if the R has reported last smoking cigars with marijuana in them more than 30 days ago, it could logically be inferred that all past month cigar use had to involve use of cigars that did not have marijuana in them.
BLNTNOMJ was skipped, even though the R reported use of blunts in the past 30 days (BLNTREC = 1). However, the R had reported lifetime use of cigars in the tobacco module but reported not smoking cigars in the past 30 days.	Rather than assign a code of 99 (LEGITIMATE SKIP), BLNTNOMJ was set to a value of 14, where 14 = USED BLUNTS PAST 30 DAYS/DIDN'T USE CIGARS PAST 30 DAYS.
BLNTNOMJ was skipped, even though the R reported use of blunts in the past 30 days (BLNTREC = 1). However, the R had reported in the tobacco module that he or she had never smoked part or all of a cigar.	Rather than assign a code of 99, BLNTNOMJ was set to a value of 24, where 24 = USED BLUNTS PAST 30 DAYS/NEVER USED CIGARS.

Table B.16 Edits in the Blunts Module Other than Those Pertaining to Consistency Checks for Incidence (continued)

Issue	Edits Implemented
BLNTNOMJ was skipped because the cigar recency variable CIGARREC or the blunts recency variable BLNTREC had missing values or "indefinite" values (i.e., used at some point in the past 12 months, or used at some point in the lifetime).	BLNTNOMJ was left as 98 (blank).
Question BL07 (corresponding to BLNTNOMJ) was answered but was not answered as "yes," and the R had not smoked blunts in the past 30 days.	BLNTNOMJ was set to a value of 3, where 3 = Yes LOGICALLY ASSIGNED. This edit affected cases where BL07 had been answered (but not as "yes"), and the Rs confirmed in BLNT30C2 that they did not use marijuana in the past 30 days.
The age at first use (AFU) was inconsistent with the R's current age despite the R being prompted to change the AFU.	The final age was accepted as the standard, and BLNTAGE was set to bad data.
BLNTAGE had a value of "don't know" (DK) or "refused" (REF), including situations where this assignment has been made from the consistency check data.	If the questions pertaining to BLNTYFU and BLNTMFU were skipped because the R answered the AFU as DK or REF, the DK or REF value from BLNTAGE was propagated onto BLNTYFU and BLNTMFU. This edit was designed to indicate the reason that BLNTYFU and BLNTMFU had been skipped. In addition, because the R may have first used blunts within 1 year of his or her current age, the month- and year-of-first-use (MFU and YFU) questions may have been relevant to the R. If the R had answered the MFU and YFU questions but BLNTAGE had a final code for DK or REF (i.e., due to a consistency check response), data in BLNTYFU and BLNTMFU were overwritten with the corresponding DK or REF value from BLNTAGE. Retaining the MFU and YFU data in this situation would imply that the R first smoked blunts within 1 year of his or her current age.
The MFU has been skipped because the R answered the YFU as DK or REF.	The DK or REF value from BLNTYFU was propagated onto BLNTMFU. That is, if the R did not know in what year he or she first used blunts, it was assumed that the R would not know the month either. Similarly, a refusal to answer the YFU was interpreted to be a blanket refusal to answer the month as well as the year.
The R had missing data for the MFU. However, one of the following occurred: • The YFU was the current calendar year, and the R was interviewed in January. • The R first used at his or her current age, the R first used in the current calendar year, and the R's most recent birthday occurred in the interview month. • The R first used at his or her current age, the R first used in the prior calendar year, and the R's most recent birthday occurred in December.	The MFU could be logically inferred, as indicated below. • The MFU was logically inferred to be January of the interview year. That would be the only month in which the R could have initiated use of blunts in the current year. • The MFU was logically inferred to be the interview month. If the R first used in the current calendar year and attained his or her current age in the interview month, the R logically could not have initiated use of blunts in any month other than the interview month. • The MFU was logically inferred to be December. If the R first used blunts in the prior calendar year and attained his or her current age in December of that year, the R logically could not have initiated use in any month other than December of the prior calendar year.

Table B.17 Edits Pertaining to the Special Topics Module

Issue	Edits Implemented
The respondent (R) reported never having been arrested or answered the lifetime arrest question as "don't know" or "refused" but reported being on probation or parole in the past 12 months.	The R was logically inferred to have been arrested at least once in his or her lifetime (i.e., BOOKED = 3). The rationale for this edit was that someone could not be on probation or parole without first having been arrested for a crime. The skipped variables pertaining to arrests in the past 12 months retained a value of blank.
The R reported being arrested in the past 12 months, did not report being arrested for a specific crime in that period, but reported being arrested for this crime as "some other offense."	The R was logically inferred to have been arrested for that crime. No further editing was done to the affirmative answer where the R reported being arrested for "some other offense" (BKOTH). Similarly, no further editing was done to the "OTHER, Specify" variable (BKOTHOFF) that indicated the crime for which the R was arrested (see Section 2.4).
The R reported being arrested at least once in the past 12 months and answered all specific past year arrest questions as "no." However, the R specified an offense in the series of follow-up questions SP03R, SP03RSP, SP03S, and SP03SSP.	The variable for "some other offense" (BKOTH) was logically inferred to be "yes." A code of 3 (i.e., Yes LOGICALLY ASSIGNED) was assigned to BKOTH. The offense that was specified in SP03RSP or SP03SSP was assigned to the edited "OTHER, Specify" variable BKOTHOFF.
The R reported being arrested at least once in the past 12 months but answered all specific past year arrest questions as "no" and reported nothing in the "OTHER, Specify" variable BKOTHOFF to support the indication of being arrested.	The response was retained to indicate that the R had been arrested in the past 12 months. A code of 5 (i.e., Offense unknown LOGICALLY ASSIGNED) was assigned to the "some other offense" variable (BKOTH).
The R reported being arrested for every offense in the past 12 months that was asked about in the module. (For youths aged 12 to 17, that included reports of being arrested for possession of tobacco; this question was skipped for adults.)	The edits differed, depending on what Rs specified for their "other" offense: • If a valid "other" offense was not specified, the entire series of past year offense variables was assigned a bad data code. • If the R gave a valid response for some other offense for which he or she was arrested in the past 12 months, the data were retained to indicate that the R was arrested for this other offense. However, the variables pertaining to arrests for all other offenses were set to bad data. • For adults, the variable pertaining to arrests for possession of tobacco (BKPOSTOB) continued to be assigned a legitimate skip code.
The R reported being arrested only one time in the past 12 months, did not report being arrested for some other offense (BKOTH = 2), but reported being arrested for every other offense in that same period.	Not including BKOTH or its associated "OTHER, Specify" variable (BKOTHOFF), the variables pertaining to arrests for specific offenses in the past 12 months were assigned a bad data code. For adults, the BKPOSTOB variable continued to be assigned a legitimate skip code.
The R reported being arrested 80 or more times in the past 12 months.	The variable pertaining to the number of arrests in the past 12 months (NOBOOKYR) was set to bad data.
The R had alternating "yes/no" or "no/yes" patterns to all questions about arrests for specific offenses in the past 12 months (e.g., SP03a = 1, SP03b = 2, SP03c = 1).	All variables pertaining to arrests for specific offenses in the past 12 months were set to bad data.

Table B.17 Edits Pertaining to the Special Topics Module (continued)

Issue	Edits Implemented
The R was asked questions about driving under the influence of alcohol or illegal drugs solely because the R originally reported past year use of one or more psychotherapeutics (i.e., pain relievers, tranquilizers, stimulants, or sedatives). However, the R was logically inferred to be a lifetime nonuser of these psychotherapeutics because the only reported lifetime use involved over-the-counter (OTC) drugs.	Any data in the substance use and driving variables (DRVALDR, DRVAONLY, and DRVDONLY) were replaced with codes of 81 (i.e., NEVER USED ALCOHOL OR DRUGS Logically assigned).
The R was asked questions about driving under the influence of alcohol, but the alcohol recency variable ALCREC had been set to bad data.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of alcohol (DRVAONLY) were set to bad data.
The R was routed into questions about driving under the influence of alcohol and illegal drugs in combination and about driving under the influence of illegal drugs, but (a) the only drug that the R definitely used in the past 12 months was alcohol (i.e., after all editing had been done to the core recency-of-use variables for alcohol and other drugs), and (b) it could not be determined that the R was not a past year user of all of the other drugs.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of illegal drugs (DRVDONLY) were set to bad data.
The R had not used alcohol in the past 12 months and was routed into the question about driving under the influence of illegal drugs solely because of psychotherapeutic use that turned out to be limited to OTC use. In addition, one or more other drug recency-of-use variables were ambiguous with respect to past year use, so it could not be determined whether the R did or did not use other illegal drugs.	The edited variable (DRVDONLY) was set to bad data.
All core drug recency variables that had triggered respondents being asked questions about driving under the influence of drugs in the past 12 months had been set to bad data.	The edited variables pertaining to driving under the influence of alcohol and illegal drugs in combination (DRVALDR) and driving under the influence of illegal drugs (DRVDONLY) were set to bad data.

Table B.18 Edits in the Prior Substance Use Module Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables

Issue	Edits Implemented
The respondent (R) indicated in the final verification check (e.g., LUCG08 for cigarettes) that the age of last use (ALU) based on the new month and year of last use (MLU and YLU) was correct (e.g., LUCG08 = 4). The computer-assisted interviewing (CAI) program updated the ALU (such as AGELSTCG for cigarettes) with the value of the age calculated from the MLU and YLU (referred to subsequently as the MYRLST age, such as MYRLSTCG, for cigarettes). However, the new value for the ALU indicated that the R was more than 1 year younger than his or her current age at the time the R last used the drug (e.g., the R was 16, reported last use of the drug at age 15, but then confirmed an MLU and YLU that meant the R was 14 when the R last used the drug). Had the R initially reported this ALU, the R would not have been routed to the MLU and YLU questions.	The updated value was retained for the ALU (e.g., last use at age 14 for a 16-year-old R in this example). Based on this updated ALU, it was logically inferred that the R should have skipped the MLU and YLU items. A code of 9989 was assigned to the YLU variable (e.g., CIGYLU for cigarettes), and a code of 89 was assigned to the MLU variable (e.g., CIGMLU).
The final verification check (e.g., LUCG08 for cigarettes) was skipped because the R entered revised data for the MLU and YLU that made them consistent with the ALU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (e.g., LUCG08) was skipped because the R entered a new MLU that was the same as the R's birth month.	The new MLU could be consistent with the ALU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the ALU, MLU, and YLU, as long as the revised MLU and YLU were potentially consistent with the ALU. However, the MLU and YLU were set to bad data if they could never be consistent with the ALU. Suppose, for example, that a hypothetical R was born in June 1994, was interviewed in March 2012 (age 17 at the time of the interview), reported last use of cigarettes at age 17, and initially reported last use in May 2011. Last use in May 2011 would have meant that the R was 16 when he or she last smoked cigarettes because the R's 17th birthday was not until June 2011. If the R changed the month and year to June 2011, that could be consistent with last use at age 17, if the use occurred after the R's birthday. However, if the R changed the month and year to June 2010, it would never be possible for the R to have last used at age 17 and also to have last used in June 2010. In this latter situation, CIGMLU and CIGYLU would be set to bad data.
The consistency check was triggered between the ALU and the MYRLST age. However, the R answered the first consistency check (e.g., LUCG05 for cigarettes, regarding whether the value in the MYRLST age was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The value for the ALU was retained, but the MLU and YLU were set to bad data.

Table B.18 Edits in the Prior Substance Use Module Pertaining to Consistency Checks for Age, Year, and Month of Last Use Variables (continued)

Issue	Edits Implemented
The final verification check (e.g., LUCG08 for cigarettes) was skipped because the R entered the same values for the MLU and YLU that triggered the inconsistency with the ALU in the first place.	The value for the ALU was retained, but the MLU and YLU were set to bad data.
The R entered a new MLU or YLU that differed from what the R previously reported. The MYRLST age based on the revised MLU and YLU still mismatched the ALU, but the R indicated in the final verification check that this MYRLST age was correct.	No editing was done in this situation. The CAI program automatically updated the ALU to be consistent with the updated values reported for the MLU and YLU.
The consistency check was triggered between the ALU and the MYRLST age calculated from the MLU and YLU. However, the R answered the first consistency check (e.g., LUCG05 for cigarettes) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The ALU was retained, but the MLU and YLU were set to bad data.
The R entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, the R failed to resolve the inconsistency between the ALU and the MYRLST age. The R also reported either in the first verification check that the MYRLST age was not correct (e.g., LUCG05 = 6 for cigarettes) or reported in the second verification check that the ALU was correct (e.g., LUCG06 = 2 for cigarettes).	No editing was done to the ALU. The following edits were implemented for the MLU and YLU: • The default edit was to set the MLU and the YLU to bad data. As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MLU and YLU were assigned the code that corresponded to the answer in the final verification check.
The R entered values in the consistency checks for the MLU and YLU that again yielded a nonmissing MYRLST age based on the MLU and YLU. However, the R failed to resolve the inconsistency between the ALU and the MYRLST age. The R also reported in the second verification check that neither the ALU nor original MYRLST age was correct (e.g., LUCG06 = 3 for cigarettes).	The following edits were implemented for the ALU, MLU, and YLU: • The default edit was to set the ALU, MLU, and YLU to bad data. As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the ALU, MLU, and YLU were assigned the code that corresponded to the answer in the final verification check.
The R triggered the initial consistency check between the ALU, MLU, and YLU and reported that the MYRLST age calculated from the MLU and YLU was correct (e.g., LUCG05 = 4). However, the R answered the second consistency check (e.g., LUCG06 for cigarettes) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the ALU, MLU, and YLU.	The relevant codes for "don't know" or "refused" were assigned to the ALU, MLU, and YLU. The rationale for this edit is that conclusive information did not exist regarding whether the ALU indicated the R's correct age when he or she last used a drug or whether the MLU and YLU indicated the R's correct age at last use. Therefore, the ALU, MLU, and YLU all were set to missing values.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks)

Issue	Edits Implemented
The age of last use (ALU) for a given drug or behavior (MRJAGLST for marijuana, CIGDLLST for daily cigarette smoking) was greater than the respondent's (R's) current age.	The ALU was set to bad data.
The ALU was lower than the age at first use (AFU) from the core (e.g., MJAGE for marijuana). If the ALU for drugs, cigarettes, or daily cigarette use was lower than the corresponding AFU, a consistency check was triggered. The Rs were not allowed to go back and change the AFU from the core; the only way they could resolve the inconsistency was by making the ALU consistent with the AFU. Thus, the Rs may indicate that the ALU was correct as reported, or they could enter a new ALU that still was inconsistent with the AFU.	The ALU was set to bad data, even if the R reported in the consistency check that "yes," what the R had previously reported for the ALU was correct. Thus, as a *partial* exception to the principle of not editing across modules (Section 2.4.1), core AFU data to edit the noncore ALU, but not vice versa.
The AFU from the core (e.g., MJAGE for marijuana) had a missing value, but the ALU (e.g., MRJAGLST) was defined. That included situations in which the AFU was set to bad data as part of the "flag and impute" edits (see Section 6.2.2.4).	No editing was done to the ALU, as long as it was consistent with the R's age (see above). In addition, no editing was done to the ALU if the imputed AFU (e.g., IRMJAGE for marijuana) was given a value greater than the ALU. (To preserve consistency of the imputed AFU data with imputed data from prior years, imputation of the core AFU variables in 2012 did not take into account data from the corresponding ALU variables.)
The edited core recency (e.g., CIGREC for cigarettes) indicated that the last use was more than 12 months ago, but the ALU indicated last use at the R's current age. The answer to the ALU would suggest use in the past 12 months. (For cigarettes, either CIGAGLST or CIGDLLST could be inconsistent with the recency.)	No editing was done to the ALU. However, a standard codebook footnote was included for these variables to alert analysts to the fact that values in this noncore module could be inconsistent with values from core modules.
The core recency was set to an "indefinite" value of 8 or 9 (used at some point in the past 12 months or used at some point in the lifetime, respectively). These indefinite recency values could be imputed to past month use. If the unedited recency in the corresponding core module indicated use more than 30 days ago, the R would be routed to the prior substance use module and could provide data for the ALU. The latter would suggest that the R is not a past month user, even though the final imputed recency might indicate that the R is a past month user.	No editing was done to the ALU. The standard codebook footnote described above also applied to this issue.
For Rs who last used drugs more than 30 days ago but within the past 12 months, the number of days that the R could have used a drug (e.g., marijuana) in the past 12 months based on MRJAGLST was less than the number of days in the edited or imputed 12-month frequency for marijuana (MJYRTOT or IRMJFY, respectively).	No editing was done to MRJAGLST. The standard codebook footnote described above also applied to this issue.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The ALU variable had a missing value, but the corresponding AFU indicated that the R first used (or started smoking cigarettes daily) at his or her current age.	• The ALU was set to equal the AFU. Logically, last use cannot be any later than the R's current age. • As part of this same edit, if the year of first use (YFU) indicated that the R first used in the year of the interview, then the R was logically inferred to have last used in the current year.
The year of last use (YLU) had a missing value. In addition, the R reported first using a drug (or first smoking cigarettes daily) at an age that was 1 year younger than his or her current age. The R also reported first using the drug (or first smoking cigarettes daily) in the current year.	No editing was done to the ALU, but the YLU was set to equal the YFU. Logically, if the first use was in the current year, then the last use had to be in the current year as well.
The month of last use (MLU) had a missing value, but the corresponding month of first use (MFU, or first daily cigarette use) was in the calendar month prior to in the interview month.	The MLU was set to equal the MFU. Logically, if the R was routed to the prior substance use module because he or she had not used a given drug in the past 30 days, the R also had to have last used that drug in the month prior to the interview month.
The R was interviewed in January and reported first use of a drug (or first daily cigarette use) in December of the previous year.	If the YLU had a missing value or had a value that was inconsistent with the YFU, then the YLU was set to equal the YFU. Concurrently, the MLU was set to equal the MFU under similar constraints.
The ALU had a missing value, but the calculated age of last use variable (MYR; e.g., MYRLSTMJ for marijuana) from the YLU, MLU, and birth date was consistent with the AFU and the R's current age. In addition, none of the above edits had been applied.	The ALU was set equal to the calculated age of last use (MYR).
The R gave a valid value for the YLU and reported last use in the same month when he or she had a birthday, or else the R answered the MLU question as "don't know" or "refused."	Two possible values for the ALU were calculated, based on the YLU that the R reported and the R's birth year. If neither of these two possible values matched the ALU that the R reported, then the YLU and MLU were set to bad data.
The R reported an ALU (and also may have reported values for the YLU and MLU) for when he or she last used any hallucinogen. However, the ALU, YLU, and MLU variables were skipped for LSD, PCP, or Ecstasy because the R had used only LSD, PCP, or Ecstasy, respectively.	Values from the ALU, YLU, and MLU for any hallucinogen (HALAGLST, HALYLU, and HALMLU, respectively) were transferred to the edited variables that had been skipped for LSD, PCP, or Ecstasy. For example, if the R was a lifetime user only of Ecstasy, data from HALAGLST, HALYLU, and HALMLU were transferred to the corresponding variables ECSAGLST, ECSYLU, and ECSMLU for Ecstasy.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The R reported an ALU (and also may have reported values for the YLU and MLU) for when he or she last used OxyContin® in the core pain relievers module or methamphetamine in the core stimulants module, respectively. However, the ALU, YLU, and MLU variables were skipped for any pain relievers or any stimulants because the R was a lifetime user of only these drugs in the respective modules.	Values from the ALU, YLU, and MLU for OxyContin® (OXYAGLST, OXYCYLU, and OXYCMLU) or methamphetamine (MTHAGLST, METHYLU, and METHMLU) were transferred to the corresponding edited variables that had been skipped for any pain relievers or any stimulants. For example, if the only pain reliever that the R had ever used nonmedically was OxyContin®, data from OXYAGLST, OXYCYLU, and OXYCMLU were transferred to the corresponding variables ANLAGLST, ANALYLU, and ANALMLU for any pain relievers. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module (see the text for further explanation).
The R reported last using any cocaine and crack cocaine in the same year. However, the R also reported last using any cocaine in a month that was earlier than the month when the R reported *first* using crack cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to bad data. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, this edit for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module (see the text for further explanation).
The ALU for crack cocaine was later than the corresponding ALU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The ALU for cocaine was set to equal the ALU for crack cocaine. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, this edit for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The ALU for any cocaine had a missing value. However, the R reported last using crack cocaine at his or her current age. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	• The ALU for any cocaine was set to equal the ALU for crack cocaine. Logically, if the R last used crack at his or her current age, then that also had to be the age at which the R last used any cocaine. • As part of this same edit, if the crack YLU variable indicated that the R last used crack in the current year, then the cocaine YLU was set to the current year. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The YLU for crack cocaine was later than the YLU for any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The YLU for any cocaine was set to equal the YLU for crack cocaine. Values in the MLU for crack cocaine also were carried over to the MLU for any cocaine. Similar edits were implemented for the YLU and MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The YLU for cocaine had a missing value. However, the R also indicated that he or she last used crack cocaine at an age that was 1 year younger than his or her current age. The R also reported last using crack cocaine in the current year. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	No editing was done to the ALU for any cocaine, but the YLU for any cocaine was set to the current year. Logically, if the last use of crack cocaine was in the current year, then the last use of any cocaine had to be in the current year as well. Similar edits were implemented for the YLU and MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The R last used any cocaine and crack cocaine at the same age and in the same year. However, the R reported last using crack cocaine in a month that was later than what the R reported for last use of any cocaine. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to equal the MLU for crack cocaine. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The cocaine MLU had a missing value, but the R reported last using crack cocaine in the calendar month prior to the interview month. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The MLU for any cocaine was set to the calendar month prior to the interview month. Logically, if the R was asked the questions about any cocaine and crack cocaine in the prior substance use module because he or she had not used these drugs in the past 30 days, the R also had to have last used any cocaine in the month prior to the interview month. Similar edits were implemented for the MLU variables for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The ALU for any cocaine had a missing value, but the calculated value in the MYR variable for crack cocaine was consistent with both the crack ALU and the R's current age. In addition, none of the above edits was applied. This issue also could occur for daily cigarette use (relative to any cigarette use), OxyContin® use (relative to any pain reliever use), and methamphetamine use reported in the core stimulants module (relative to any stimulant use).	The ALU for any cocaine was set to equal the MYR value for crack cocaine. Similar edits were implemented for the ALUs for any cigarette use, any pain reliever use, or any stimulant use when this issue applied in their respective sections of the prior substance use module. However, these edits for methamphetamine and stimulants did not apply when last use of methamphetamine was based on methamphetamine use identified in the noncore special drugs module.
The R reported last using any hallucinogen and LSD, PCP, or Ecstasy in the same year. However, the R also reported last using any hallucinogens in a month that was earlier than the month when the R reported *first* using LSD, PCP, or Ecstasy.	The MLU for any hallucinogens was set to bad data.
At least one of the ALUs for LSD, PCP, and Ecstasy was greater than the ALU for any hallucinogen.	• The hallucinogen ALU was set to the latest ALU from the ALUs for LSD, PCP, or Ecstasy. • As part of this same edit, if a given ALU for LSD, PCP, or Ecstasy was moved over to the ALU for any hallucinogens and the corresponding YLU for LSD, PCP, or Ecstasy was later than the YLU for any hallucinogen, then the later YLU was moved over to the YLU for any hallucinogen. If two or all three of these drugs had the same latest ALU value relative to the hallucinogen ALU, then the latest YLU from these drugs was moved over to the YLU for any hallucinogen. • If the YLU for LSD, PCP, or Ecstasy was moved over to the YLU for any hallucinogen and the hallucinogen MLU had a nonmissing value that was earlier than the corresponding MLU for LSD, PCP, or Ecstasy, then the MLU for any hallucinogen was set to the latest MLU from LSD, PCP, or Ecstasy.
The ALU for any hallucinogens had a missing value, but one or more ALUs for LSD, PCP, or Ecstasy indicated that the R last used at his or her current age.	• The hallucinogen ALU was set to be equal to the R's current age. • As part of this same edit, if a given YLU for LSD, PCP, or Ecstasy indicated that the R last used in the current year, then the hallucinogen YLU was set to the current year.
The ALU for any hallucinogens had a missing value, and one or more ALUs for LSD, PCP, or Ecstasy indicated that the R last used at a year younger than his or her current age. At least one YLU for LSD, PCP, or Ecstasy indicated that the R last used any of these drugs in the current year.	The hallucinogen YLU was set to the current year. No editing was done to the hallucinogen ALU.

Table B.19 Edits in the Prior Substance Use Module Pertaining to the Age, Year, and Month of Last Use (Other than for Consistency Checks) (continued)

Issue	Edits Implemented
The MLU for any hallucinogens had a missing value, but the R reported last using LSD, PCP, or Ecstasy in the calendar month prior to the interview month.	The MLU for any hallucinogens was set to the calendar month prior to the interview month. Logically, if the R was asked the questions about any hallucinogens and LSD, PCP, or Ecstasy because he or she had not used these drugs in the past 30 days, the R also had to have last used any hallucinogens in the month prior to the interview month.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables

Issue	Edits Implemented
Respondents (Rs) last used a particular psychotherapeutic drug in the past 30 days. They reported getting the drug "in some other way" in the past 30 days and specified a source that corresponded to one of the available response options for obtaining the drug in the past 30 days.	If Rs did not report getting the drug from that particular source in the past 30 days, the corresponding variable was assigned a code of 3. For example, if Rs did not report that they got pain relievers from a friend or relative for free in the past 30 days but they reported this in the "OTHER, Specify" variable ANLOTHSP (corresponding to question LU28SP), then the 30-day variable for getting pain relievers from a friend or relative for free (ANLFRFRE) was assigned a code of 3. In this example, a code of 3 meant, "Response entered LOGICALLY ASSIGNED (from ANLOTHSP)." Similar edits were implemented if Rs reported that they obtained the drug from a friend or relative for free in the past 30 days, the friend or relative obtained the drug "in some other way," and the source of the friend or relative's drug corresponded to one of the available response options for how the friend or relative obtained medication that the R used nonmedically in the past 30 days.
Rs last used a particular drug in the past 30 days. They reported getting the drug in a specific way the last time they used it nonmedically that they did not report for the past 30 days.	Logically, if Rs last used a drug in the past 30 days, the way that they got the drug the last time they used it nonmedically also applied to the past 30 days. The corresponding past month variable was assigned a code of 5. For example, if Rs did not report that they got pain relievers from a friend or relative for free in the past 30 days but they reported in question LU28 that they got the last pain relievers they misused in this way, then ANLFRFRE was assigned a code of 5. In this example, a code of 5 meant, "Response entered LOGICALLY ASSIGNED (from ANLGTLAS)." Similar edits were implemented if Rs reported that they obtained the drug from a friend or relative for free in the past 30 days, and the Rs also reported that the friend or relative obtained the drug that the Rs last used in a way that the Rs did not report for the 30-day items on how the friend or relative obtained the drug.
Rs were logically inferred to have gotten medication from friends or relatives for free in the past 30 days (e.g., ANLFRFRE = 3 or 5). Consequently, Rs were not asked how the friends or relatives obtained the medication.	The edited variables pertaining to how the friends or relatives obtained the medication retained a code of 98 (i.e., blank). If the Rs had reported directly that they obtained a psychotherapeutic drug from a friend or relative for free (instead of the response being logically inferred), they would have been asked the subsequent questions about how the friend or relative obtained the drug.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables (continued)

Issue	Edits Implemented
Rs reported that they got the last drug they used nonmedically in "some other way," but they specified getting it in one of the ways that they had been asked about previously.	The code from the "OTHER, Specify" variable plus a value of 10 was added to the variable for how Rs obtained the drug the last time they used it nonmedically (i.e., "bump" by 10). For example, if Rs reported getting pain relievers in some other way when they last used them nonmedically (LU28 = 10), but they specified getting them from a friend or relative for free (i.e., ANLGTOSP = 5), then ANLGTLAS was coded as 15. Similar edits were implemented if Rs reported that friends or relatives obtained a drug that the Rs last used in "some other way" and then the Rs specified one of the ways they had asked about previously. For example, if ANLFFOSP was coded as 5 (i.e., the friend or relative got the pain reliever from another friend or relative for free), then ANLFFLAS (how the friend or relative obtained the pain reliever that the R last used nonmedically) was coded as 15.
Rs were logically inferred to have gotten the last medication they used nonmedically from friends or relatives for free (e.g., ANLGTLAS = 15 for pain relievers; see above). Consequently, Rs were not asked how the friends or relatives obtained the medication that the Rs last used.	The corresponding edited variable pertaining to how the friends or relatives obtained the medication that the Rs last used retained a code of 98 (i.e., blank).
Rs were logically inferred to have gotten a drug from one of the sources they were asked about for the last time they used it nonmedically. The Rs also used the drug in the past 30 days and did not report getting the drug from that source in the past 30 days.	The corresponding 30-day variable was assigned a code of 5. For example, if ANLGTLAS = 15 because getting pain relievers from a friend or relative for free was specified as "some other way" that Rs got the pain relievers the last time they used them nonmedically, then ANLFRFRE was coded as 5, if it was not already coded as 1 or 3 (e.g., if Rs had not previously specified getting pain relievers from a friend or relative for free as "some other way" they got them in the past 30 days). Similar edits were in place for the variables pertaining to how friends or relatives obtained a drug that Rs used nonmedically in the past 30 days. For example, if Rs used prescription pain relievers nonmedically in the past 30 days, got the pain relievers they used the last time from friends or relatives for free, and the friends or relatives were logically inferred to have gotten the pain relievers in a particular way, then the friends or relatives were logically inferred to have gotten the pain relievers in this way for pain relievers that the Rs used nonmedically in the past 30 days.
Rs reported that they or friends/relatives got a drug in the past 30 days or the last time in "some other way." However, the Rs explicitly specified that the drug they or friends/relatives obtained was an over-the-counter (OTC) drug.	No editing was done when Rs specified that they had obtained OTCs.

Table B.20 Edits in the Prior Substance Use Module Pertaining to the Source of Psychotherapeutics Variables (continued)

Issue	Edits Implemented
Rs last used a particular drug in the past 30 days. They reported getting the drug in "some other way" the last time they used it nonmedically, and they did not report getting the drug in some other way in the past 30 days.	The variable for obtaining the drug in some other way in the past 30 days was assigned a code of 5. The "OTHER, Specify" response for how Rs got the drug the last time they used it nonmedically also was transferred over to the "OTHER, Specify" variable for the past 30 days (provided that Rs had not specified that they obtained OTCs). For example, if Rs reported that they got pain relievers in some other way the last time they used them nonmedically and they specified getting them from a friend, relative, or at home, but they did not specify whether they got them for free, bought them, or took them without asking (ANLGTOSP = 11), then the 30-day variable ANLOTHWY (for "some other way") was assigned a code of 5, and the corresponding "OTHER, Specify" variable ANLOTHSP was coded as 11 as well. Similar edits were implemented if Rs reported that they got a drug from a friend or relative in the past 30 days. In particular, the "other" response for how friends or relatives obtained the drug that the Rs last used nonmedically was transferred to the corresponding "OTHER, Specify" variable for how friends or relatives obtained the drug that Rs used nonmedically in the past 30 days (e.g., ANLFFOSP for pain relievers), provided that the R did not specify that the friend or relative had obtained OTCs.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana

Issue	Edits Implemented
All age-at-first-use (AFU) variables for these drugs had codes of 997, indicating the respondent (R) refused to answer the AFU questions for these drugs, or the R refused to answer the lifetime use question.	A code of 97 (REFUSED) was assigned to each sequence-of-use variable USEALCG, USEMJCG, USEALMJ, USEACM, or USENEXT that was blank. Otherwise, if the R was routed to a given question for some reason, a bad data code was assigned to the edited variable.
One or both AFUs for the drugs of interest had missing values, but the corresponding sequence-of-use variable had a valid value. For example, the AFU variables AGE1STAL and AGE1STCG that were stored by the computer-assisted interviewing (CAI) program had equal values while the interview was in progress, such that the R was asked USEALCG. In editing of the core data, however, the cigarette AFU variable CIGTRY could have been set to bad data because of an inconsistency with the cigarette recency CIGREC. Similarly, the alcohol AFU variable ALCTRY could have been set to bad data because of an inconsistency with the alcohol variable ALCREC.	A code was assigned that was equal to the original value of the sequence variable plus 10 (i.e., "bump" by 10). This edit preserved data indicating that the R was routed to a particular item and how the R answered, but it would alert analysts that one or both of the AFUs that routed the R to the item was questionable. For example, if question LU22, corresponding to USEALCG, was coded as 1, indicating that the R started using alcohol before using cigarettes, but the AFU for alcohol or the AFU for cigarettes had missing values (e.g., bad data), then USEALCG was assigned a code of 11.
The R initiated use of a given pair of drugs (e.g., alcohol and cigarettes) within 1 year of his or her current age. However, the sequence variables were not consistent with the *edited* year-of-first-use (YFU) and month-of-first-use (MFU) variables from the core modules. For example, the R reported in question LU22 that he or she first used alcohol before using cigarettes. However, the edited YFU and MFU data from the tobacco and alcohol modules indicated that the R first used cigarettes in an earlier year or month than when the R first used alcohol.	Where edited YFU and MFU data existed for a given pair of drugs, the edits logically inferred a sequence of use that was consistent with the core YFU and MFU data. For example, if the R answered question LU22 as 1 (corresponding to USEALCG), that would indicate that the R used alcohol before using cigarettes. However, if the cigarette YFU and MFU data indicated use of cigarettes before alcohol, USEALCG was assigned a code of 4, where 4 = Cigarettes LOGICALLY ASSIGNED. Similarly, if the R answered LU22 as 2, indicating that the R used cigarettes before using alcohol, but the alcohol YFU and MFU data indicated use of alcohol before cigarettes, then a code of 3 was assigned to USEALCG, where 3 = Alcohol LOGICALLY ASSIGNED. These edits also were performed when USEALCG has missing values (e.g., "don't know" [DK] or "refused" [REF]), but YFU and MFU data from the core could be used to infer a nonmissing value in USEALCG. These principles also applied to edits of USEMJCG and USALMJ. The rationale for these edits was that the Rs themselves provided data in the core modules that indicated the sequence with which they used these drugs.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana (continued)

Issue	Edits Implemented
Values of 1 or 2 in question LU26 (corresponding to USENEXT) had different meanings, depending on how the R answered question LU25 (corresponding to USEACM).	USENEXT was recoded to parallel the levels in USEACM (1 = Alcohol; 2 = Cigarettes; 3 = Marijuana). • USEACM = 1 indicated that the R first used alcohol before using cigarettes or marijuana. When USEACM = 1, a value of 1 in LU26 indicated that the R used cigarettes next. A value of 2 in LU26 indicated that the R used marijuana next. Therefore, when USEACM = 1, USENEXT was coded as 2 when LU26 = 1 and was coded as 3 when LU26 = 2. • USEACM = 2 indicated that the R first used cigarettes before using alcohol or marijuana. When USEACM = 2, a value of 1 in LU26 indicated that the R used alcohol next. A value of 2 in LU26 indicated that the R used marijuana next. Therefore, when USEACM = 2, USENEXT was coded as 1 when LU26 = 1 and was coded as 3 when LU26 = 2. • USEACM = 3 indicated that the R first used marijuana before using cigarettes or alcohol. When USEACM = 3, a value of 1 in LU26 indicated that the R used alcohol next. A value of 2 in LU26 indicated that the R used cigarettes next. Therefore, when USEACM = 3, USENEXT was coded as 1 when LU26 = 1 and was coded as 2 when LU26 = 2.
Data existed in USEACM (LU25), but one or more AFUs had been set to bad data (e.g., because of an inconsistency in the core modules between the AFU and the recency).	Values other than those for DK, RE, or bad data were bumped by 10 in USEACM and USENEXT. As was the case for USEALCG, USEMJCG, and USEALMJ, this edit preserved data in USEACM and USENEXT while making analysts aware that there is a potential issue with the data.
Question LU25 (corresponding to USEACM) was answered as 1, indicating that the R first used alcohol. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for cigarettes and marijuana were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used cigarettes before using marijuana or alcohol, the edits assigned a code of 5 to USEACM, where 5 = Cigarettes LOGICALLY ASSIGNED. For example, if the respective YFU variables for cigarettes, alcohol, and marijuana, CIGYFU, ALCYFU, and MJYFU, all were defined and CIGYFU was earlier than the other two (but, by definition, the R first used all three substances at the same age), it could logically be inferred that cigarette use occurred first. • Similarly, if it could be determined unambiguously that the R first used marijuana before using cigarettes or alcohol, the edits assigned a code of 6 to USEACM, where 6 = Marijuana LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either marijuana or cigarettes before using alcohol but incidence data were not fully defined for all three substances, USEACM was set to bad data.

Table B.21 Edits in the Prior Substance Use Module Pertaining to the Sequence of Initiation for Alcohol, Cigarettes, and Marijuana (continued)

Issue	Edits Implemented
Question LU25 (corresponding to USEACM) was answered as 2, indicating that the R first used cigarettes. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for alcohol and marijuana were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used alcohol before using marijuana or cigarettes, the edits assigned a code of 4 to USEACM, where 4 = Alcohol LOGICALLY ASSIGNED. • Similarly, if it could be determined unambiguously that the R first used marijuana before using cigarettes or alcohol, the edits assigned a code of 6 to USEACM, where 6 = Marijuana LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either alcohol or marijuana before using cigarettes but incidence data were not fully defined for all three substances, USEACM was set to bad data.
Question LU25 (corresponding to USEACM) was answered as 3, indicating that the R first used marijuana. However, this response was inconsistent with one or more YFU or MFU variables from the core modules for cigarettes, alcohol, or marijuana.	The core YFU and MFU data for cigarettes and alcohol were checked for indications that the R used either of these drugs first. • If it could be determined unambiguously that the R used alcohol before using marijuana or cigarettes, the edits assigned a code of 4 to USEACM, where 4 = Alcohol LOGICALLY ASSIGNED. • Similarly, if it could be determined unambiguously that the R first used cigarettes before using alcohol or marijuana, the edits assigned a code of 5 to USEACM, where 5 = Cigarettes LOGICALLY ASSIGNED. Otherwise, if there was some indication that the R used either alcohol or cigarettes before using marijuana but incidence data were not fully defined for all three substances, USEACM was set to bad data.
USEACM had been assigned codes of 4, 5, 6, or 85 (bad data).	To edit USENEXT, a sequence of logic was followed similar to that outlined above for USEACM. For example, if USEACM had been set to a value of 4 (used alcohol first) and the core YFU and MFU data for cigarettes and marijuana indicated that the R used cigarettes next, set USENEXT was set to a value of 5 (Cigarettes LOGICALLY ASSIGNED), if USENEXT did not already indicate that the R used cigarettes next. If USEACM had been set to bad data and USENEXT was not blank, USENEXT also was set to bad data.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004

Issue	Edits Implemented
The respondent's (R's) only report(s) of drug use in the core drug modules that routed the R into question TX01 about lifetime substance treatment had been set to bad data as part of the core drug editing.	Nonblank values in the edited variables pertaining to receipt of substance treatment were replaced with bad data codes.
Responses to the questions on the receipt of treatment in the past 12 months and the last time that the R received treatment were inconsistent (e.g., if the R reported that he or she did not receive treatment in the past 12 months but subsequently reported last receiving treatment during that period).	The edits favored responses that indicated more recent receipt of treatment: • If an R responded affirmatively that he or she had received treatment in the past 12 months but reported last receiving treatment "more than 12 months ago," the edits logically inferred that the R last received treatment at some point in the past 12 months (i.e., TXLASREC = 8). • If an R reported that he or she did not receive treatment in the past 12 months but reported last receiving treatment in the past 12 months, the edits logically inferred that the R had received treatment in that period (i.e., TXYREVER = 3).
The question on the receipt of treatment in the past 12 months had missing data (e.g., a response of "don't know" or "refused"), but the question on the last time that the R received treatment did not. Alternatively, the question on the last time that the R received treatment had missing data, but the question on receipt of treatment in the past 12 months did not.	Where possible, data were used to replace the missing value with a nonmissing value. Suppose, for example, that the R did not know or refused to report whether he or she had received treatment in the past 12 months. • If the R reported last receiving treatment in this period, the ambiguous response was replaced with a value to indicate that the R had received treatment in this period (i.e., TXYREVER = 3). • If the R reported last receiving treatment more than 12 months ago, it was logically inferred that the question about receipt of any treatment in the past 12 months should have been answered as "no" (i.e., TXYREVER = 4).
The question about the most recent receipt of treatment had missing data.	Data from the past year treatment variable (TXYREVER) or the lifetime treatment variable (TXEVER) were used to replace missing values in the edited treatment recency variable TXLASREC. • If the R had received treatment in the past 12 months (TXYREVER = 1), TXLASREC was assigned a code of 8 to indicate treatment at some point in the past 12 months. • If the R had not received treatment in the past 12 months (TXYREVER = 2), TXLASREC was assigned a code of 13 to indicate that the last treatment episode was more than 12 months ago. • If the question about treatment in the past 12 months (TX02) was answered as "don't know" or "refused" and the R did not report currently being in treatment in question TX07, TXLASREC was assigned a code of 9 to indicate treatment at some point in the R's lifetime. (The R had to have answered the lifetime treatment question TX01 as "yes" in order to have been routed to TX02.)

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported currently being in treatment in question TX07, so the question about the most recent time that the R had been in treatment was skipped.	The edited variable corresponding to question TX07 (TXRCVNOW) continued to be coded as 1 (i.e., "yes"). Instead of a legitimate skip code being assigned, the edited treatment recency variable (TXLASREC) was assigned a code of 7, where 7 = Still in treatment LOGICALLY ASSIGNED. A code of 21 (still in treatment) also was assigned to the treatment outcome variable TXLTYOUT.
The R reported currently being in treatment in question TX07 but did not know or refused to report in question TX02 whether he or she had received treatment in the past 12 months.	The R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3).
The R was routed to question TX07 but did not know or refused to report whether he or she was still in treatment.	Data from the treatment outcome variable (TXLTYOUT) and treatment recency variable (TXLASREC) were used to replace missing values in the edited variable TXRCVNOW, corresponding to TX07. • If the R had received treatment in the past 30 days (TXLASREC = 1) and reported still being in treatment (TXLTYOUT = 1 or 21), TXRCVNOW was assigned a code of 3 to indicate that the R logically was still in treatment. • Otherwise, if the R definitely had not received treatment in the past 30 days, TXRCVNOW was assigned a code of 4 to indicate that the R logically was no longer in treatment.
The R reported that he or she was not currently in treatment (TXRCVNOW = 2 or 4), but the R reported still being in treatment when asked about the outcome of the last treatment episode.	The treatment outcome variable (TXLTYOUT) was assigned a bad data code.
The R specified receiving treatment for an over-the-counter (OTC) psychotherapeutic medication (e.g., aspirin).	This information on OTC drugs was not used to infer treatment for any of the psychotherapeutic drugs because the questions about receipt of treatment for psychotherapeutic drugs referred specifically to treatment for prescription-type medications (i.e., not to OTCs).
The R did not report receiving treatment for a particular drug during his or her last (or current) treatment episode, but treatment for this drug was specified as treatment for "some other drug." In the case of the psychotherapeutics, the "other" drug specified was not an OTC drug.	The R was inferred to have received (or be receiving) treatment for the use of that drug. A code of 3 (Yes LOGICALLY ASSIGNED) was assigned to the corresponding edited drug variable. For example, Rs who did not report receiving treatment for prescription stimulants but reported receiving treatment for street stimulants were considered to qualify as having received treatment for prescription-type stimulants (i.e., those that were not available as OTCs, which would include street drugs). In this example, the edited variable TXLTYSTM would be assigned a code of 3.
The R did not report receiving treatment for a particular drug during his or her last (or current) treatment episode but indicated that this drug was the primary drug for which he or she last received treatment (or was currently receiving treatment).	The R was inferred to have received (or be receiving) treatment for the use of that drug.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment only for alcohol in the past 12 months, but questions about treatment for specific drugs during the last or current treatment episode had missing values (i.e., "don't know," "refused," bad data, or blank).	The R was logically inferred not to have received treatment for that drug during the last or current episode. The missing value in that drug's variable was replaced with a special code of 4 (No LOGICALLY ASSIGNED).
The R was routed to questions about the last or current treatment episode but did not have any indication of treatment for any of the drugs that he or she ever used.	The following edits were implemented: • If the R reported receiving treatment only for alcohol in the past 12 months, a special logically inferred "yes" code of 5 was assigned to the variable for alcohol treatment during the last treatment episode (TXLTYALC). • If the R reported receiving treatment only for drugs other than alcohol in the past 12 months, a special code of 5 was assigned to the "some other drug" variable (TXLTYSOD) to indicate that the drug for which the R received treatment was unknown. • Otherwise, a special code of 7 was assigned to TXLTYSOD, the "some other drug" variable, to indicate that treatment for alcohol or other drugs was unknown. (Prior to 2002, these edits required Rs to have denied receiving treatment for all drugs they had ever used. Since 2002, the above edits also have been implemented if Rs did not report treatment for any specific drugs, and missing data existed in the questions about treatment for specific drugs.)
The R was logically inferred to have received treatment for alcohol during the last or current treatment episode (TXLTYALC = 5), and question TX36 about treatment for any other drug (TXLTYSOD) was answered as "no."	The edited "OTHER, Specify" variables TXLTYA through TXLTYE were assigned legitimate skip codes.
The R refused to report in question TX36 whether he or she received treatment for any other drug.	The refusal was propagated onto the edited "OTHER, Specify" variables TXLTYA through TXLTYE. Since 2002, this edit has been implemented regardless of whether the R had reported treatment for at least one drug in questions TX26 through TX35. (Prior to 2002, this edit required at least one response of "yes" in TX26 through TX35.)
The R reported treatment only for "some other drug," but the only substances specified were tobacco products (i.e., cigarettes, chewing tobacco, snuff, cigars, pipe tobacco).	The variables specifying treatment for tobacco products were assigned bad data codes. In addition, other variables pertaining to the last (or current) treatment episode were assigned bad data codes if the items had been answered. The following variables were affected: TXLTYMN (i.e., main place where the R was last treated); TXLTYOUT (i.e., outcome of the last treatment episode); variables beginning with TXPY (i.e., payment sources for the last [or current] treatment episode); and TXLTYDUR (i.e., length of the last or current treatment). The rationale for these edits was that anything pertaining to the last treatment (e.g., payment sources for the last or current treatment) would logically be assumed to pertain to treatment only for tobacco.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
If the R reported in the alcohol module that he or she never used alcohol (AL01 = 2), the R would be skipped out of question TX26, pertaining to receipt of treatment for alcohol. However, the R also could report in question TX03 that he or she received treatment for "alcohol only" or "alcohol and drugs" in the past 12 months.	The edited variable pertaining to receipt of alcohol treatment during the last or current episode (TXLTYALC) retained a code of 98 (blank).
Question TX37 pertaining to the main drug for which the R last received (or was currently receiving) treatment was skipped because the R reported receipt of treatment for only one drug during the last or current treatment episode. That includes situations in which the only drug for which the R reported receiving treatment was "some other drug."	The edited variable TXLTYPRM was assigned a legitimate skip code, provided that none of the edited variables about treatment for alcohol through prescription sedatives (TXLTYALC through TXLTYSED) had a code of 98 (blank). Otherwise, TXLTYPRM retained a code of 98.
The R reported being in treatment for 366 days in the past 12 months.	The edited variable TXLTYDUR was edited to 365 days.
The length of time that the R reported currently being in treatment or being in treatment the last time translated to a number of years greater than the R's age.	The edited variable TXLTYDUR was assigned a bad data code.
The R reported receiving treatment in the past 12 months and reported receiving treatment in the past 12 months for alcohol only or drugs only. However, this response was inconsistent with the responses to questions on the drugs for which the R was treated (or was being treated) during the last (or current) treatment episode. For example, the R reported being treated in the past 12 months only for alcohol but reported last being treated for use of one or more other drugs.	Logically, the last or current treatment episode would fall within the 12-month period prior to the interview. Therefore, the variable pertaining to receipt of treatment for alcohol, other drugs, or both in the past 12 months (TXYRADG) was edited as follows: • If the R originally indicated treatment for alcohol only (i.e., a code of 1 in question TX03), with treatment for other drugs also having been indicated during the last episode, a special code of 11 was assigned to TXYRADG. • If the R originally indicated treatment for drugs only (i.e., a code of 2 in question TX03), with treatment for alcohol also having been indicated during the last episode, a special code of 12 was assigned to TXYRADG. The edits were done in this manner because the subsequent fill pattern for specific locations where the R received treatment in the past 12 months was based on the R's original answer for receipt of treatment only for alcohol, only for other drugs, or both.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment in the past 12 months but did not know or refused to report whether he or she received treatment only for alcohol, only for other drugs, or for both. However, data were provided on the drugs for which the R was treated during his or her last (or current) treatment episode.	Data on the drugs for which the R was last treated (or was currently being treated) were used to indicate the *minimum* for which the R could have been treated in the past 12 months: • If the R indicated last (or currently) being treated for alcohol but did not indicate treatment for other drugs during the last (or current) treatment episode, it was possible to infer in TXYRADG that the R was at least treated for alcohol in the past 12 months in TXYRADG (but the R also may have been treated for other drugs at some point during that period). A special code of 4 was assigned to TXYRADG. • If the R indicated last (or currently) being treated for one or more drugs other than alcohol but did not indicate treatment for alcohol, it was possible to infer in TXYRADG that the R was at least treated for drugs other than alcohol in the past 12 months. A special code of 5 was assigned to TXYRADG. • If the R reported last (or currently) being treated both for alcohol and for other drugs, it was possible to infer in TXYRADG that the R was treated for both alcohol and other drugs in the past 12 months. A special code of 6 was assigned to TXYRADG.
The R reported receiving treatment in the past 12 months and did not report receiving treatment in a particular location in the past 12 months, but this location was specified as treatment in "some other place" in the past 12 months.	The R was logically inferred to have received treatment in that location in the past 12 months. A code of 3 (Yes LOGICALLY ASSIGNED) was given to the edited treatment location variable in this situation.
The R reported receiving treatment in the past 12 months (or was inferred to have received treatment in the past 12 months) and did not report receiving treatment in a particular location that he or she subsequently reported was the main place that he or she received treatment the last time (or the main place where he or she was currently receiving treatment).	The R was logically inferred to have received treatment in that location in the past 12 months. A special logically assigned "yes" code of 5 was assigned to indicate that the affirmative response came from the data on the main location where the R last received (or was currently receiving) treatment. If the R reported that the main location where he or she received treatment was "some other place" and specified a valid response in question TX25SP, that "OTHER, Specify" response also was moved over to the "OTHER, Specify" variable TXYROTSP pertaining to treatment locations in the past 12 months.
The R reported receiving treatment in the past 12 months but answered "no" to every item about particular locations for treatment in that period, including "some other place."	The edited variable pertaining to "some other place" (TXYRSOP) was assigned a special code of 7 to indicate that the treatment location was unknown.
The R reported receiving treatment in the past 12 months and did not initially indicate receiving treatment in a hospital emergency room in that period. However, the R subsequently reported receiving treatment in the past 12 months in an emergency room for use of marijuana, cocaine, heroin, LSD, PCP, or methamphetamine.	The variable that did not indicate treatment in an emergency room (TXYRTXER) was edited to infer that the R had received treatment in that location in the past 12 months.

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R reported receiving treatment in the past 12 months in every specific location that was asked about (i.e., except for treatment in "some other place").	No editing was done if the R reported being or having been in treatment for 15 days or more. If the R reported being or having been in treatment for fewer than 15 days, however, responses of "yes" in the entire list of edited past year treatment location variables were replaced with bad data codes. If treatment in "some other place" also was reported, the edited variable TXYRSOP was assigned a bad data code. In the variable TXYROTSP (i.e., the other treatment location that was specified), any responses were replaced with bad data codes. If the R also reported that he or she was still in treatment (TX07 = 1), the edited variable TXRCVNOW also was assigned a bad data code.
Rs could report still being in treatment in question TX07 but may report that they received treatment only in jail in the past 12 months.	When Rs reported receiving treatment only in jail in the past 12 months, they were logically inferred not to be currently in treatment (TXRCVNOW = 4). This edit also applied when Rs reported that the main treatment location was jail, and this was the only specific location reported for treatment in the past 12 months.
The R did not report a particular payment source for his or her last episode of treatment but specified this payment source as "some other source."	The R was inferred to have used that particular payment source for treatment. A code of 3 (Yes LOGICALLY ASSIGNED) was assigned to the edited variable for that payment source.
The R answered all items about payment sources for treatment as "no," including the item indicating that the last treatment was free.	A special code of 5 was assigned to the edited "some other source" variable (TXPYSOS) to indicate that the payment source was unknown.
The R reported that every specific payment source that was asked about paid for his or her last episode of treatment (i.e., except for "some other source" and free payment, the latter of which would have been skipped).	All source of payment variables that the R had answered as "yes" were assigned a bad data code. That included situations in which "some other source" of payment also was reported. In the variable TXPYSP (i.e., the other payment source that was specified), any responses were replaced with bad data codes.
The R reported *all* of the following: • receipt of treatment in every specific location in the past 12 months (i.e., except for treatment in "some other place"), and • payment of the last treatment by every specific payment source (i.e., except for "some other source" and free treatment).	When this specific pattern occurred, data from additional variables also were assumed to be questionable. Responses entered for the following variables were replaced with bad data codes: TXYRADG (i.e., treatment for alcohol, drugs, or both in the past 12 months); TXYRVSER (treatment in an emergency room for marijuana, cocaine, heroin, LSD, PCP, or methamphetamine in the past 12 months); TXYRNMER (number of times the R visited an emergency room for treatment of the above drugs); TXLTYMN (the main place the R received treatment the last time); drugs that the R was asked about for the last treatment episode (including the main drug for which the R received treatment, if applicable); and TXLTYDUR (length of time in treatment currently or the last time).

Table B.22 Edits Pertaining to the Receipt of Substance Treatment Variables That Existed Prior to 2004 (continued)

Issue	Edits Implemented
The R indicated that "some other source" paid for the last treatment, but then specified that this treatment was free.	If no other payment source was indicated, then it was logically inferred that the R's last treatment was free (i.e., TXPYFRE = 3). Otherwise, if one or more payment sources had been indicated previously (e.g., private health insurance, the R's own funds), then it was inferred that "some other source" had not paid for the last treatment. In this situation, the response of free treatment that had been specified also was wiped out in the edited "OTHER, Specify" variable (TXPYSP).
The R reported in question TX44 that the only treatment he or she received in the past 12 months was for detoxification (or answered TX44 as "don't know" or "refused"), but the R also reported attending self-help groups in the past 12 months. Self-help groups typically are not places where people go to receive detoxification.	The response was accepted that the R received treatment in a self-help group in the past 12 months, and the R was logically inferred to have received treatment other than detoxification in that period. The edited variable corresponding to question TX44 (TXYRDTXO) was assigned a code of 4 (No LOGICALLY ASSIGNED).

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004

Issue	Edits Implemented
The respondent (R) reported receiving substance treatment in the past 12 months (i.e., TXYREVER = 1).	The following lifetime treatment variables were assigned legitimate skip codes: TXALONEV, TXALONAG, TXDRONEV, TXDRONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG. These variables corresponded to questions TX45, TX45A, TX46, TX46A, TX47, TX47A, TX48, and TX48A, respectively.
The R was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3). Consequently, the R was routed into the lifetime treatment questions. Had the R answered the past year treatment question TX02 as "yes," however, the R would have been skipped out of the lifetime treatment questions.	Responses of 1 and 2 (i.e., "yes" and "no," respectively) in TXALONEV, TXDRONEV, TXALCDEV, and TXDRGAEV were bumped by a value of 10 (i.e., set to values of 11 or 12). No editing was done to these variables if they had values of 94 ("don't know") or 97 ("refused").
The R received treatment in his or her lifetime but not in the past 12 months, and the R never used alcohol (i.e., ALCEVER = 2). The R also reported using at least one other drug in the marijuana through sedatives modules or in the special drugs module for methamphetamine.	The following variables pertaining to lifetime alcohol treatment were assigned legitimate skip codes: TXALONEV, TXALONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG.
The R received treatment in his or her lifetime but not in the past 12 months, and the R indicated unambiguously in the marijuana through sedatives modules or in the special drugs module for methamphetamine that he or she never used any of these drugs. The R also reported using alcohol.	All the variables pertaining to drugs are skip filled, namely, TXDRONEV, TXALCDEV, TXDRGAEV and the age variables TXDRONAG, TXALCDAG, and TXDRGAAG.
The R reported first receiving substance treatment for alcohol or other drugs (or both) at an age that was later than his or her current age.	Consistent with standard editing practice, any ages for first receipt of treatment that were inconsistent with the R's current age were set to bad data.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R reported first receiving treatment for alcohol or other drugs at an age that was earlier than the age-at-first-use (AFU) data from the core drug use modules or the special drugs module for methamphetamine. • For alcohol, the R reported first receiving treatment for alcohol at an age that earlier than ALCTRY, the AFU for alcohol from the core alcohol module. • For other drugs, the R reported first receiving treatment for his or her use of drugs at an age that was earlier than the minimum AFU for the drugs that the R reported using.	No editing was done to the age at first treatment data that were inconsistent with the AFU data from the core substance use modules. However, flags were created to indicate the degree of deviation from (a) ALCTRY (TXFGALAG), (b) the minimum AFU for other drugs (TXFGDGAG), and (c) the minimum AFU from both ALCTRY and the AFUs for other drugs (TXFGADAG). Values in these flags had the following meanings: 0 = First treatment age was consistent with minimum AFU 1 = First treatment age differs from minimum AFU by 1 year 2 = First treatment age differs from minimum AFU by 2 years 3 = First treatment age differs from minimum AFU by 3 or 4 years 4 = First treatment age differs from minimum AFU by 5 or more years. The flag for both alcohol and drugs (TXFGADAG) was created from the values in TXFGALAG and TXFGDGAG. The final value in TXFGADAG was chosen according to whatever value in TXFGALAG and TXFGDGAG indicated the greatest inconsistency between treatment age and core AFU data. If both TXFGALAG and TXFGDGAG were blank, then TXFGADAG was set to blank as well.
The R could report receiving treatment for alcohol in the past 12 months and could give an age when he or she first received alcohol treatment. However, the R previously reported in the core alcohol module that he or she never used alcohol. Similarly, the R could report receiving treatment for other drugs in the past 12 months and could give an age when he or she first received drug treatment. However, the R previously reported never using drugs in the core marijuana through sedatives modules or in the special drugs module for methamphetamine.	Again, no editing was done to the inconsistent treatment age data. However, the flag variables mentioned above were assigned a code of 5 when this situation occurred. This code of 5 had the following meaning: 5 = Never used alcohol in core but reported an alcohol treatment age 5 = Never used drugs in core but reported a drug treatment age. If TXFGALAG or TXFGDGAG had a value of 5, then TXFGADAG also was assigned a value of 5, where 5 = Any lifetime nonuse with a related treatment age.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R answered the lifetime treatment question TX01 (edited variable TXEVER) as "don't know" or "refused."	The following edits were implemented for the lifetime treatment variables TXALONEV through TXDRGAAG according to what the R reported in the core substance use modules for lifetime use (or nonuse) of alcohol or drugs. • If the R was a lifetime user only of alcohol (i.e., had definitely never used other drugs that were covered in the core modules), the lifetime variables pertaining to treatment for drugs (TXDRONEV, TXDRONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG) were assigned legitimate skip codes; even if the R had reported lifetime receipt of treatment, these variables still would have been skipped because the R never used any of the drugs covered in the core module. In addition, if TXEVER was refused, that refusal was propagated to TXALONEV and TXALONAG. • If the R was a lifetime user of other drugs but had never used alcohol, the lifetime variables pertaining to treatment for alcohol (TXALONEV, TXALONAG, TXALCDEV, TXALCDAG, TXDRGAEV, and TXDRGAAG) were assigned legitimate skip codes; even if the R had reported lifetime receipt of treatment, these variables still would have been skipped because the R never used alcohol. In addition, if TXEVER was refused, that refusal was propagated to TXDRONEV and TXDRONAG.
The R had the following pattern in his or her data: 1. Lifetime use of alcohol (i.e., ALCEVER = 1). 2. Initially skipped out of past year treatment questions but was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3; see Table B.22). 3. Never used marijuana through sedatives, or methamphetamine (based on answers from special drugs). 4. Question TX45 (edited variable TXALONEV) was answered as "no," meaning that the R had never received treatment or counseling for alcohol. Logically, the R had to have received treatment for something, and alcohol was the only substance that the R reported using.	TXALONEV was bumped by 20, such that TXALONEV would show a value of 22. This was done instead of inferring that the R received alcohol treatment. Nevertheless, a value of 22 in TXALONEV would alert analysts to an inconsistency in the data and give them the option of deciding how to handle these cases in an analysis.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The R had the following pattern in his or her data: 1. Lifetime use of at least one drug in the marijuana through sedatives modules or methamphetamine use from the special drugs module. 2. Initially skipped out of past year treatment questions but was logically inferred to have received treatment in the past 12 months (i.e., TXYREVER = 3). 3. Never used alcohol. 4. Question TX46 (edited variable TXDGONEV) was answered as "no," meaning that the R had never received treatment or counseling for drugs other than alcohol. Logically, the R had to have received treatment for something, and illicit drugs (but not alcohol) were the only substances that the R reported using.	TXDRONEV was bumped by 20, such that TXDRONEV would show a value of 22. This was done instead of inferring that the R received drug treatment. Nevertheless, a value of 22 in TXDRONEV would alert analysts to an inconsistency in the data and give them the option of deciding how to handle these cases in an analysis.
The R reported lifetime use of both alcohol and other drugs. However, both TXALCDEV = 2 and TXDRGAEV = 2, meaning that the R never got treatment for either alcohol or drugs. Logically, if the R answered the lifetime treatment question TX01 (edited variable TXEVER) as "yes," the R had to have gotten treatment for something.	Both TXALCDEV and TXDRGAEV were bumped by 20, such that the edited value would be 22. Again, this would give analysts the option of deciding how to handle these cases in an analysis.
A lifetime lead variable (i.e., TXALONEV, TXDRONEV, TXALCDEV, and TXDRGAEV) was coded as 2 (i.e., "no") or was coded as 12 (based on the second issue described in this table).	The corresponding age-at-first-treatment variables TXALONAG, TXDRONAG, TXALCDAG, and TXDRGAAG were assigned legitimate skip codes.
A past year lead variable (i.e., TXYALDRG, TXYDRALC) was coded as 2 (i.e., "no").	The corresponding age-at-first-treatment variables TXYALDAG and TXYDRAAG were assigned legitimate skip codes.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
The past year treatment variable TXYRADG (corresponding to question TX03) indicated that the R was treated in the past year only for alcohol or only for drugs.	If TXYRADG = 1 (i.e., treatment in the past year only for alcohol), the following variables were assigned legitimate skip codes: TXYDRAGE (age at first treatment for drugs, if the R had been treated in the past year only for drugs), TXYDRALC (whether the R ever was treated for alcohol use, if the R had ever used alcohol and had been treated in the past year only for drugs), TXYDRAAG (the treatment age variable corresponding to TXYDRALC), TXYADAAG (age at first treatment for alcohol, if the R had been treated in the past year for both alcohol and drugs), and TXYADDAG (age at first treatment for drugs, if the R had been treated in the past year for both alcohol and drugs). Similarly, if TXYRADG = 2 (i.e., treatment in the past year only for drugs), the following variables were assigned legitimate skip codes: TXYALAGE (age at first treatment for alcohol, if the R had been treated in the past year only for drugs), TXYALDRG (whether the R ever had been treated for drug use, if the R had ever used marijuana through sedatives and had been treated in the past year only for alcohol), TXYALDAG (the treatment age variable corresponding to TXYALDRG), TXYADAAG (age at first treatment for alcohol, if the R had been treated in the past year for both alcohol and drugs), and TXYADDAG (age at first treatment for drugs, if the R had been treated in the past year for both alcohol and drugs).
TXYRADG indicated that the R was treated in the past year for both alcohol and drugs, including situations in which the R was logically inferred to have received treatment for both alcohol and drugs in the past year.	The following variables were assigned legitimate skip codes: TXYALAGE, TXYALDRG, TXYALDAG, TXYDRAGE, TXYDRALC, and TXYDRAAG. If TXYRADG = 6 (i.e., logically inferred to have received treatment for both alcohol and drugs in the past year; see Table B.22), the skipped variables TXYADAAG and TXYADDAG retained codes of blank.
TXYRADG had been assigned a code of 4 (i.e., received treatment for alcohol in the past year but treatment for drugs during this period was unknown; see Table B.22).	The following variables pertaining to receipt of treatment only for drugs were assigned legitimate skip codes: TXYDRAGE, TXYDRALC, and TXYDRAAG. The skipped variables pertaining to treatment for alcohol in the past year, with or without treatment for drugs (i.e., TXYALAGE, TXYALDRG, TXYALDAG, TXYADAAG, and TXYADDAG), retained codes of blank.
TXYRADG had been assigned a code of 5 (i.e., received treatment for alcohol in the past year but treatment for drugs during this period was unknown; see Table B.22).	The following variables pertaining to receipt of treatment only for alcohol were assigned legitimate skip codes: TXYALAGE, TXYALDRG, and TXYALDAG. The skipped variables pertaining to treatment for drugs in the past year, with or without treatment for alcohol (i.e., TXYDRAGE, TXYDRALC, TXYDRAAG, TXYADAAG, and TXYADDAG), retained codes of blank.

Table B.23 Edits Pertaining to the Receipt of Substance Treatment Variables That Were Added in 2004 (continued)

Issue	Edits Implemented
TXYRADG had been assigned a code of 11 (i.e., the R had reported receiving treatment only for alcohol in the past 12 months but some treatment for drugs also was indicated; see Table B.22). TXYALDRG did not indicate that the R had ever received treatment for his or her use of drugs.	TXYALDRG was assigned a code of 3 (i.e., Yes LOGICALLY ASSIGNED).
TXYRADG had been assigned a code of 12 (i.e., the R had reported receiving treatment only for drugs in the past 12 months but some treatment for alcohol also was indicated; see Table B.22). TXYDRALC did not indicate that the R had ever received treatment for his or her use of alcohol.	TXYDRALC was assigned a code of 3 (i.e., Yes LOGICALLY ASSIGNED).
TXYRADG had been set to bad data.	Any nonblank values in variables that were dependent on TXYRADG were set to bad data. This edit was relevant to the following questions (corresponding edited variables shown in parentheses): TX49 (TXYALAGE), TX49A (TXYALDRG), TX49B (TXYALDAG), TX50 (TXYDRAGE), TX50A (TXYDRALC), TX50B (TXYDRAAG), TX51 (TXYADAAG), and TX51A (TXYADDAG).
TXYRADG was answered as "don't know" or "refused."	All of the variables TXYALAGE through TXYADDAG that had been skipped retained codes of blank.

Table B.24 Edits Pertaining to the Perceived Need for Treatment Variables

Issue	Edits Implemented
The only indication(s) of lifetime drug use that routed the respondent (R) into the substance treatment questions had been set to bad data because only over-the-counter (OTC) drug use had been reported in the core.	Nonblank values in the edited variables pertaining to perceived need for substance treatment were replaced with bad data codes.
The R specified the need for treatment for an OTC psychotherapeutic medication (e.g., aspirin).	This information on OTC drugs was not used to infer need for treatment for any of the psychotherapeutic drugs because the questions about perceived need for treatment for psychotherapeutic drugs referred specifically to prescription-type medications (i.e., and not OTCs).
The R did not report needing treatment for a particular drug in the past 12 months, but need for treatment for this drug was specified as a treatment need for "some other drug." In the case of the psychotherapeutics, the other drug specified was not an OTC drug.	The R was inferred to perceive the need for treatment for the use of that drug. For example, Rs who did not report needing treatment for prescription stimulants but reported needing treatment for street stimulants were considered to qualify as perceiving the need for treatment for prescription-type stimulants (i.e., those that were not available over the counter, which would include street drugs). The edited variable NDTXSTMR was assigned a code of 3 (Yes LOGICALLY ASSIGNED). This code of 3 could be edited further, as discussed below.
The R reported needing treatment in the past 12 months for the use of alcohol or other drugs, but questions about the perceived need for treatment for all specific drugs that the R had ever used were answered as "no."	A special code was assigned to the "some other drug" variable (NDTXSOD) to indicate that the specific drug for which the R thought that he or she needed treatment was unknown.
Question TX10, pertaining to the perceived need for additional treatment, is an "enter all that apply" type of question. That is, Rs could report needing additional treatment for more than one drug shown in the list in TX10. However, Rs could report needing additional treatment for drugs that they had reported never using in the corresponding core module (e.g., reported never using heroin but reported needing additional treatment for heroin). In contrast, Rs would not get asked questions TX11 through TX21 (regarding perceived need for treatment for specific drugs) unless they were lifetime users of a particular drug.	No editing was done when this pattern occurred. Consequently, these noncore data would be inconsistent with the core data.

Table B.24 Edits Pertaining to the Perceived Need for Treatment Variables (continued)

Issue	Edits Implemented
The R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3). Because the R did not originally answer question TX02 as "yes," the computer-assisted interviewing (CAI) program routed the Rs to questions about whether they thought they needed treatment for their use of alcohol or specific drugs (i.e., question TX08 and questions TX11 through TX22).	The following edits were done when TXYREVER = 3: • If a question was originally answered as "yes," then the corresponding edited variable was assigned a code of 11 (Yes [TXYREVER = 3]). For example, if the R reported needing treatment for alcohol or other drugs (TX08 = 1), then the edited variable NDTXNEDR was assigned a code of 11. Similarly, if the R reported needing treatment for a specific drug (e.g., prescription stimulants), then the edited variable (e.g., NDTXSTMR) was assigned a code of 11. • If a question was originally answered as "no," then the corresponding edited variable was assigned a code of 12 (No [TXYREVER = 3]). For example, if TX08 had been answered as "no" (TX08 = 2), then NDTXNEDR was assigned a code of 12. (If NDTXNEDR was set to 12, then subsequent variables continued to be assigned legitimate skip codes.) Similarly, if a question about the need for treatment for a specific drug had been answered as" no," then the edited variable was assigned a code of 12. • If the R was inferred to perceive the need for treatment for a drug based on "OTHER, Specify" data, the edited variable was assigned a code of 13. Suppose, for example, that NDTXSTMR had already been coded as 3 because the R had specified prescription-type stimulants as "some other drug" for which the R needed treatment (but question TX19 had not been answered as "yes"). If the R was logically inferred to have received treatment in the past 12 months (TXYREVER = 3), then NDTXSTMR was subsequently coded as 13 (Yes LOGICALLY ASSIGNED [TXYREVER = 3]). • If the R was a lifetime nonuser of a drug, the edits continued to assign a legitimate skip code. For example, if the R had never used prescription-type stimulants, then NDTXSTMR continued to receive a code of 99 when TXYREVER = 3. • Codes for any reasons that respondents reported for why they did not get treatment were bumped by 10; the resulting codes were 11 or 13. The rationale for these edits was that Rs would not have been asked questions about their perceived need for treatment for alcohol or specific other drugs if they had originally reported that they received treatment in the past 12 months. The above edits were done to conserve respondents' answers, as opposed to wiping out the data.
The R reported making an effort to get treatment (question TX22 answered as "yes"), but the R reported not needing treatment for every specific drug that he or she was asked about.	The edited variable NDTXEFTR was assigned a code of 11. The same edits described above for other variables that applied when TXYREVER = 3 also were performed when NDTXEFTR was assigned a code of 11 due to this issue.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables

Issue	Edits Implemented
Respondents (Rs) did not choose an outpatient treatment location from the list of locations in question ADMT14, but that location was specified as a source of outpatient mental health treatment in the past 12 months in AUOPYRSP.	The edited variable corresponding to receipt of outpatient treatment at that location was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report receiving outpatient mental health counseling at the office of a private therapist, reported receiving outpatient counseling in "some other place," and specified something to indicate that he or she received counseling from a private therapist, the edited variable AUOPTHER was assigned a code of 3.
Rs reported receiving mental health services in every inpatient or outpatient location in a list.	For inpatient treatment, all of the variables corresponding to the service locations were set to bad data, including the numbers of nights that Rs reported spending at these various inpatient treatment settings. For outpatient treatment, the edits depended on what Rs specified for the "other" outpatient location where they received mental health treatment. • If a valid "other" outpatient location was not specified, the entire series of outpatient variables (including the reported numbers of visits) was set to bad data. • If the R reported a valid "other" outpatient location where he or she received mental health services in the past 12 months, the data were retained to indicate that the R received services in this location. However, the remaining variables pertaining to receipt of outpatient mental health treatment were set to bad data.
Rs reported at least one of the following: (a) they stayed overnight as an inpatient for mental health treatment in a particular type of facility for 365 or 366 days in the past 12 months, or (b) they stayed overnight as an inpatient in more than one type of facility, and the total number of nights that they stayed as inpatients summed to 365 or more.	If Rs reported inpatient treatment in a particular location for 366 days in the past 12 months, the corresponding edited variable (e.g., AUNMPSYH for the number of nights hospitalized in a psychiatric hospital) was reset to 365. No other editing was done when these patterns occurred.
Rs did not choose a payment source for their mental health treatment but subsequently indicated that this was (or would be) the principal payment source.	The edited payment source variable was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report that private health insurance paid or would pay for outpatient treatment but then reported that private insurance was (or would be) the principal source of payment, the edited variable AUPOPINS (private health insurance paid/will pay for any outpatient mental health treatment) was assigned a code of 3.
Rs reported a specific source of payment for their services but also reported that "No one paid because the treatment was free."	No editing was done because these responses were not necessarily inconsistent. Rs could have received services in more than one setting or from more than one provider, with some services being free and other services requiring payment.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables (continued)

Issue	Edits Implemented
Rs did not report a specific reason in question ADMT27 or ADMT27A for why they did not receive mental health treatment in the past 12 months, but they specified this as "some other reason."	The edited variable associated with that particular reason for not receiving mental health treatment was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if Rs specified that they did not get mental health treatment because they could not afford the cost and they had not chosen that reason in question ADMT27, the edited variable AUUNCOST (no mental health treatment because the R could not afford the cost) was assigned a code of 3. Similarly, if Rs specified that they did not get treatment because they thought they could handle the problem without treatment but had not chosen that reason in ADMT27A, the edited variable AUUNHNDL (no mental health treatment because the R thought he or she could handle the problem without treatment) was assigned a code of 3.
Rs did not choose a particular alternative service provider from the list of providers in question ADMT29b, but that provider was specified as a source of alternative mental health treatment or support in the past 12 months.	The edited variable corresponding to receipt of alternative treatment from that type of provider was assigned a code of 3 (Response entered LOGICALLY ASSIGNED). For example, if an R did not report receiving treatment or support from a chiropractor in question ADMT29b, reported receiving treatment from some other provider, and specified something to indicate that he or she received treatment or support from a chiropractor, the edited variable AUALCHIR was assigned a code of 3.
Rs did not report receiving inpatient mental health treatment, or they reported receiving inpatient treatment but not at a location listed in question ADMT02. However, the Rs also specified treatment in a particular inpatient location as some other source of "alternative" treatment in AUALOTSP.	If a specific inpatient treatment location had not been reported in AUINPSYH through AUINRESD but the Rs specified treatment in that location in AUALOTSP, the edited inpatient variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUALOTSP). In addition, if AUINPYR, pertaining to receipt of any inpatient mental health treatment in the past 12 months, was not answered as "yes," AUINPYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.
Rs did not report receiving outpatient mental health treatment, or they reported receiving outpatient treatment but not at a location listed in question ADMT14. However, the Rs also specified treatment in a particular outpatient location as some other source of "alternative" treatment in AUALOTSP.	If a specific outpatient treatment location had not been reported in AUOPMENT through AUOPDTMT but the Rs specified treatment in that location in AUALOTSP, the edited outpatient variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUALOTSP). This code of 5 was designed to allow analysts to distinguish between logical inferences based on AUOPYRSP (see above) and those based on AUALOTSP. In contrast, if the R did not report receiving mental health treatment in an outpatient medical clinic, for example, a code of 3 in AUOPCLNC would mean that the R had not reported treatment in an outpatient medical clinic but reported receiving treatment in that location as some other outpatient location, from AUOPYRSP. In addition, if AUOPTYR, pertaining to receipt of any outpatient mental health treatment in the past 12 months, was not answered as "yes," AUOPTYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.

Table B.25 Edits Pertaining to the Adult Mental Health Service Utilization Variables (continued)

Issue	Edits Implemented
Rs did not report taking medication in the past 12 months that was prescribed for a mental health condition. However, the Rs also specified that they took medication as some other source of "alternative" treatment in AUALOTSP.	The edited variable AURXYR, pertaining to taking prescribed medication in the past 12 months, was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.
Rs did not report receiving treatment from any alternative providers in the past 12 months, or they reported receiving alternative treatment but not from a provider listed in question ADMT29B. However, the Rs also specified treatment from a particular alternative provider as some other source of "outpatient" treatment in AUOPYRSP.	If treatment from a specific alternative provider had not been reported in AUALACUP through AUALMASG but the Rs specified treatment from that location in AUOPYRSP, the edited alternative treatment variable was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from AUOPYRSP). This code of 5 was designed to allow analysts to distinguish between logical inferences based on AUALOTSP (see above) and those based on AUOPYRSP. For example, if the R did not report receiving mental health treatment from a chiropractor but reported receiving outpatient treatment from a chiropractor in AUOPYRSP, the edited variable AUALCHIR (mental health treatment from a chiropractor in the past 12 months) was assigned a code of 5. In comparison, a code of 3 in AUALCHIR would mean that the R had not reported treatment from a chiropractor but reported receiving that treatment from some other provider in AUALOTSP. In addition, if AUALTYR, pertaining to receipt of any alternative mental health treatment in the past 12 months, was not answered as "yes," AUALTYR was assigned a code of 3, where 3 = Yes LOGICALLY ASSIGNED.

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module

Response Pattern	Edit
Data from the core alcohol module indicated that the respondent (R) never used alcohol.	Codes of 91 (or 991, 9991) were assigned to the edited variables in this module. This edit took precedence over situations in which questions may not have been applicable for other reason (e.g., if Rs were male).
Even if the R had used alcohol, data from remaining core modules indicated that the R had never used drugs other than alcohol.	Codes of 91 were assigned to the edited variables corresponding to questions CA08 and CA09, pertaining to alcohol use in combination with other drugs in the past 30 days (edited variables CADRKDRG through CADRKSED). The code of 91 had the following meaning in these variables: NEVER USED ALCOHOL OR NEVER USED OTHER DRUGS. Similarly, if the R had not used illicit drugs and the R had been logically inferred to be a nonuser of at least one drug, then these edited variables were assigned codes of 81, where 81 = NEVER USED ALCOHOL OR NEVER USED DRUGS Logically assigned.
Data from the core alcohol module indicated that the R had used alcohol, but definitely not in the past 30 days.	Codes of 93 (or 993, 9993) were assigned to the edited variables in this module to indicate that the R did not use alcohol in the past 30 days.
Even if the R had used alcohol in the past 30 days, data from remaining core modules indicated that the R had used some other drugs in the lifetime but not in the past 30 days.	Codes of 93 were assigned to the edited variables corresponding to questions CA08 and CA09. The code of 93 had the following meaning in these variables: DID NOT USE ALCOHOL OR DID NOT USE DRUGS IN THE PAST 30 DAYS. In addition, data were overwritten with codes of 83 (DID NOT USE ALCOHOL OR DID NOT USE DRUGS IN THE PAST 30 DAYS Logically assigned) in CADRKDRG (corresponding to question CA08) if the only reports of drug use in the past 30 days had been changed to indicate logical inference of lifetime nonuse because the only drugs that the R had ever reported using in the past 30 days were over-the-counter (OTC) psychotherapeutic drugs.
The R had used alcohol in the past 30 days (or was potentially a past month user, based on ALCREC = 8 from the core alcohol module), but the R was aged 21 or older.	The edited variables corresponding to questions CA02 through CA07SP about underage alcohol use were assigned codes of 99 (LEGITIMATE SKIP) if they were blank or otherwise were assigned codes of 89 (LEGITIMATE SKIP Logically assigned).

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

Response Pattern	Edit
The R had used alcohol in the past 30 days (or was potentially a past month user) and was aged 12 to 20.	• If the R did not indicate in question CA02B that he or she drank the alcohol in some other place the last time, then the edited variable CADROTHSP (corresponding to question CA02SP) was assigned a legitimate skip code (i.e., 99 if blank or 89 otherwise). • If the R indicated in question CA03 (edited variable CABUYFRE) that the R paid for the last alcohol that he or she drank: – Edited variables corresponding to questions CA07 (edited variable CAFREWHO) and CA07SP (edited variable CAFRESP) were assigned a legitimate skip code. – If the R paid for the last alcoholic beverage that he or she drank but the R gave money to someone else who bought it (question CA04 = 2, corresponding to the edited variable CAGVMONY), then the edited variables corresponding to questions CA05 (CABUYWHO), CA05A (CABPLACE), and CA05B (CABUNDAG) were assigned a legitimate skip code. – If the R paid for the last alcoholic beverage that he or she drank and also bought it (CAGVMONY = 1), then the edited variable corresponding to question CA06 (CAGVWHO) was assigned a legitimate skip code. In addition, if the R reported in CA05 that he or she bought the alcohol in a store, in a restaurant, in a club, or at an event (CA05 = 1), then CABUNDAGE was assigned a legitimate skip code. If the R reported in question CA05 that he or she bought the alcohol from another person (CA05 = 2), then CABPLACE was assigned a legitimate skip code. • If the R indicated in question CA03 that he or she did not pay for the last alcohol that he or she drank: – CAGVMONY, CABUYWHO, CABPLACE, CABUNDAG, and CAGVWHO were assigned a legitimate skip code. – If the R did not report in question CA07 that he or she got the last alcoholic beverage some other way, then the edited variable CAFRESP (corresponding to CA07SP) was assigned a legitimate skip code.

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

Response Pattern	Edit
The R had used alcohol in the past 30 days and also had used other drugs in that period.	• If the R reported in question CA08 (edited variable CADRKDRG) that he or she did not use other drugs at the same time or within a couple of hours of the last time when the R drank alcohol in the past 30 days, then the edited variables corresponding to the drugs in question CA09 (i.e., edited variables CADRKMRJ through CADRKSED) were assigned legitimate skip codes. • Otherwise (e.g., if CA08 was answered as "yes"), if the core data indicated that the R had not used a particular drug in the past 30 days but the R had used some other drug in the past 30 days, then the variable corresponding to the drug that the R had not used was assigned a legitimate skip code. Suppose, for example, that the R used other drugs at the same time or within a couple of hours of when the R last used alcohol and the R had used marijuana but not cocaine in the past 30 days. In this situation, CADRKCOC (i.e., use of cocaine in combination with alcohol) was assigned a legitimate skip code.
The R reported using other drugs at the same time or within a couple of hours of when he or she last used alcohol. However, question CA09 about use of alcohol in combination with specific drugs was skipped because the R reported use of only one illicit drug in the past 30 days.	The edited variable corresponding to the one illicit drug that the R had used in the past 30 days was assigned a code of 5, where 5 = Response entered LOGICALLY ASSIGNED (from skip pattern). For all other drugs that the R had definitely not used in the past 30 days, the edited variables were assigned legitimate skip codes. For example, if the R answered question CA08 as "yes," and the only drug that the R reported using in the past 30 days was marijuana, then it could be logically inferred that the R used marijuana at the same time or within a couple of hours of when he or she last used alcohol, and the edited variable CADRKMRJ was assigned a code of 5.
The R was a lifetime alcohol user but was male.	The edited variables corresponding to questions CA12 through CA14d about females' consumption of four or more drinks on a single occasion were assigned codes of 99 (or 999, 9999) if they were blank or otherwise were assigned codes of 89 (or 989, 9989).
The lifetime binge alcohol question CA10 (for consumption of five or more drinks on a single occasion) had been skipped because either of the following occurred: • The core data indicated some consumption of five or more drinks on a single occasion in the past 30 days. • Question CA01 (edited variable CADRLAST) indicated that the R had five or more drinks the last time that he or she drank in the past 30 days.	The edited variable CABNGEV (corresponding to question CA10) was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).
Question CA10 was answered as "no," indicating that the R had never had five or more drinks on a single occasion.	The edited variables CABNGAGE (corresponding to question CA11), CABNGYFU (derived from questions CA11a through CA11c), and CABNGMFU (derived from questions CA11c and CA11d) were assigned legitimate skip codes.

Table B.26 Prioritization of Specific Skip Logic Edits for the Consumption of Alcohol Module (continued)

Response Pattern	Edit
The lifetime binge alcohol question CA12 for females (for consumption of four or more drinks on a single occasion) had been skipped because any of the following occurred: • The core data indicated that the R usually had four or more drinks on those days when she drank in the past 30 days. • CADRLAST indicated that the R had four or more drinks the last time she drank in the past 30 days. • Question CA10 (edited variable CABNGEV) was answered as "yes," indicating that the R had consumed five or more drinks on a single occasion at least once in her lifetime.	The edited variable CA4FDEV (corresponding to question CA12) was set to a value of 5, where 5 = Yes LOGICALLY ASSIGNED (from skip pattern).
For female Rs, question CA12 was answered as "no," indicating that these Rs had never had four or more drinks on a single occasion. There also was no indication from CA10 that the R had ever had five or more drinks on a single occasion.	• The edited variables CA4FDAGE (corresponding to question CA14), CA4FDYFU (derived from questions CA14a through CA14C), and CA4FDMFU (derived from questions CA14C and CA14D) were assigned legitimate skip codes. • If the R had used alcohol in the past 30 days (or the most recent period of alcohol use was ambiguous and could have included the past 30 days), and there was no other indication that she had ever had four or more drinks on a single occasion, then the edited variable CA4FDDYS (corresponding to question CA13) was assigned a legitimate skip code. Otherwise, if the R was a lifetime alcohol user but definitely had not used alcohol in the past 30 days, then the edits described above took precedence, and CA4FDDYS was assigned a code of 93.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables

Issue	Edits Implemented
The respondent (R) indicated in the final verification check for the age when he or she first engaged in binge alcohol use (BACC06 for five or more drinks and WBACC06 for four or more drinks) that the age at first binge alcohol use (AFU) based on the new month and year of first use (MFU and YFU) was correct (i.e., BACC06 = 4 or WBACC06 = 4). The computer-assisted interviewing (CAI) program updated the AFU (AGE1STBA or AGE1WBA for female respondents) with the value of the age calculated from the MFU and YFU (i.e., MYR1STBA or MYR1WBA). However, the new value for the AFU indicated that the R was more than 1 year younger than his or her current age at the time the R first engaged in binge alcohol use. Had the R initially reported this AFU, the R would not have been routed to the relevant MFU and YFU questions.	The updated value was retained in the edited AFU variables CABNGAGE or CA4FDAGE. Based on the updated value for the AFU, it was logically inferred that the R should have skipped the MFU and YFU items. A code of 9989 was assigned to the relevant YFU variable (CABNGYFU or CA4FDFYU), and a code of 89 was assigned to the corresponding MFU variable (CABNGMFU or CA4FDMFU).
The final verification check (BACC06 or WBACC06) was skipped because the R entered revised data for the relevant MFU and YFU that made MFU and YFU for first binge alcohol use consistent with the AFU.	No editing was done because the R was considered to have resolved the inconsistency.
The final verification check (BACC06 or WBACC06) was skipped because the R entered a new MFU that was the same as the R's birth month.	The new MFU could be consistent with the AFU, depending on whether the use in that month occurred before or after the R's birthday. No editing was done to the AFU, MFU, and YFU, as long as the revised MFU and YFU were potentially consistent with the AFU. However, the revised MFU and YFU were set to bad data if they could never be consistent with the AFU.
The R entered a new MFU or YFU that differed from what the R previously reported. The age based on the revised MFU and YFU (updated in MYR1STBA or MYR1WBA) still mismatched the AFU, but the R indicated in the final verification check that the new value from the relevant MYR variable was correct.	No editing was done in this situation. The CAI program automatically updated the value for the AFU to be consistent with the updated values reported for the MFU and YFU.
The consistency check was triggered between the "AGE1" AFU variable stored by the CAI program (e.g., AGE1STBA for consumption of five or more drinks) and the corresponding MYR variable (e.g., MYR1STBA). However, the R answered the first consistency check (e.g., BACC03, regarding whether the value in MYR1STBA was correct) as "don't know" or "refused." The R then exited the consistency check loop without having resolved the inconsistency.	The AFU value that had been stored in the AGE1 variable (e.g., in AGE1STBA) was retained, but the corresponding MFU and YFU variables were set to bad data.
The consistency check was triggered between the AGE1 and corresponding MYR variables. However, the R entered the same values for the YFU and MFU that triggered the inconsistency with the AFU in the first place.	No editing was done to the AFU, but the YFU and MFU variables were set to bad data.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR variable. However, the R failed to resolve the inconsistency between the AGE1 variable and the updated value in the MYR variable. The R also reported either in the first verification check that the MYR was not correct (e.g., BACC03 = 6) or reported in the second verification check that the AFU from the AGE1 variable was correct (e.g., BACC04 = 2).	No editing was done to the AFU. The following edits were implemented for the MFU and YFU: • The default edit was to set the MFU and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the MFU and YFU were assigned the code that corresponded to the answer in the final verification check.
The R entered new values in the consistency checks for the MFU and YFU that again yielded a nonmissing value for the MYR. However, the R failed to resolve the inconsistency between the AGE1 variable and the updated value in the MYR variable. The R also reported in the second verification check that neither answer was correct for what was originally captured in the AGE1 and MYR variables (e.g., i.e., BACC04 = 3).	The following edits were implemented for the AFU, MFU, and YFU: • The default edit was to set the AFU, MFU, and YFU to bad data. • As an exception to the default edit, if the final verification check was answered as "don't know" or "refused," the AFU, MFU, and YFU were assigned the code that corresponded to the answer in the final verification check.
The AFU was inconsistent with the value from the MYR (based on the MFU and YFU). The R also reported that the age from the MYR was correct (e.g., BACC03 = 4). However, the R answered the second consistency check (BACC04) as "don't know" or "refused." Consequently, the R did not have an opportunity to correct the inconsistency between the AFU, MFU, and YFU.	The relevant codes for "don't know" or "refused" were assigned to the AFU, MFU, and YFU. The rationale for this edit is that conclusive information did not exist regarding whether the AFU indicated the R's correct age when he or she first engaged in binge alcohol use, or whether the MFU and YFU indicated the R's correct age at initiation of binge alcohol use. Therefore, the AFU, MFU, and YFU all were set to missing values.
The R reported first binge alcohol use at an age that was earlier than what was reported for his or her first use of any alcohol or at an age that was greater than the R's current age.	The edited AFU variable (e.g., CABNGAGE) was set to bad data. Any nonmissing values in the YFU and MFU also were set to bad data.
The R reported first binge alcohol use in a year that was earlier than what the R reported for the year in which he or she first used any alcohol.	The edited YFU variable was set to bad data. Any nonmissing values in the MFU also were set to bad data.
The R reported first binge alcohol use in the same year in which the R first used any alcohol. However, the R reported first binge alcohol use in a month that was earlier than the month in which the R first used any alcohol.	The edited MFU variable was set to bad data.
The R reported first binge alcohol use in the same month as his or her birth month. However, the R could never have first engaged in binge alcohol use at the age he or she reported and also have first engaged in binge alcohol use in the year that the R reported.	The relevant YFU and MFU variables were set to bad data.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
An edited AFU variable for binge alcohol use had a missing value. However, the R first used alcohol at his or her current age.	• The binge alcohol AFU variable with the missing value was set to the R's current age. • The corresponding editing indicator (EI) variable was set to a value of 2, where 2 = Logically assigned data. • If the YFU and MFU had been set to bad data but the unedited values that the R reported for the YFU and MFU were the same as those in ALCYFU and ALCMFU, then the relevant binge YFU and MFU variables were reset to their unedited values.
An edited YFU variable for binge alcohol use had a missing value. However, the R first used alcohol in the current survey year.	• The binge alcohol YFU variable with the missing value was set to the current year. • If the edited YFU was not the same as the unedited value, then the corresponding EI variable was set to a value of 2, where 2 = Logically assigned data. • If the MFU had a missing value and ALCMFU indicated that the R first used alcohol in the interview month, then the binge MFU was set to the same value as ALCMFU. If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (e.g., EIBNGMFU) was set to a value of 2.
An edited MFU variable for binge alcohol use had a missing value. However, the YFU data and interview date supplied information on when the R logically first engaged in binge alcohol use.	• The missing value in the MFU was replaced with the definite month in which the R logically first engaged in binge alcohol use. For example, if the R reported first use of any alcohol and first binge alcohol use in the same year and the R first used any alcohol in December of that year, then December logically was the only month in that year in which the R could have initiated binge alcohol use. • If the edited MFU value was not the same as the unedited value, then the corresponding EI variable for the MFU (e.g., EIBNGMFU) was set to a value of 2.
For females, the R reported first consumption of five or more drinks on a single occasion that was earlier than the age she reported for first consumption of four or more drinks.	• The earlier AFU from CABNGAGE was assigned to CA4FDAGE, and EI4FDAGE was set to a value of 2. • If CABNGYFU had a nonmissing value, then data from CABNGYFU and CABNGMFU also were transferred to CA4FDYFU and CA4FDMFU. • If the edited values for CA4FDYFU and CA4FDMFU differed from the corresponding unedited values, then EI4FDYFU and EI4FDMFU were assigned a value of 2.

Table B.27 Edits in the Consumption of Alcohol Module Pertaining to the Binge Alcohol Use Variables (continued)

Issue	Edits Implemented
For females, the R reported first consumption of five or more drinks on a single occasion in a year that was earlier than the year she reported for first consumption of four or more drinks.	• The earlier YFU from CABNGYFU was assigned to CA4FDYFU, and EI4FDYFU was set to a value of 2. • Data from CABNGMFU also were transferred to CA4FDMFU. • If the edited value for CA4FDMFU differed from the corresponding unedited value, then EI4FDMFU was assigned a value of 2.
For females, the R reported first consumption of five or more drinks on a single occasion and four or more drinks on a single occasion in the same year. However, the R reported first consumption of five or more drinks in a month that was earlier than what she reported for first consumption of four or more drinks.	• Data from CABNGMFU were transferred to CA4FDMFU. • If the edited value for CA4FDMFU differed from the corresponding unedited value, then EI4FDMFU was assigned a value of 2.
For females, CA4FDDYS (corresponding to question CA13) had a missing value. However, the edited variable DR5DAY from the core alcohol module indicated that these Rs had five or more drinks on a single occasion on all 30 days in the past month.	CA4FDDYS was assigned a value of 30, and EI4FDDYS was assigned a value of 2.
For females, CA4FDDYS had a nonmissing value that was less than the number of days from DR5DAY.	CA4FDDYS was assigned the value from DR5DAY, and EI4FDDYS was assigned a value of 2.
For females, the R reported that she had four or more drinks on a single occasion on 0 days in the past 30 days. However, CADRLAST or the core alcohol data indicated some consumption of four or more drinks in that period.	CA4FDDYS was set to bad data.

Appendix C: Race and Hispanic/Latino Group Alpha Codes

Appendix C: Race and Hispanic/Latino Group Alpha Codes

C.1 Introduction

For the 2012 National Survey on Drug Use and Health (NSDUH), it was not uncommon for a respondent to feel that the categories for race or Hispanicity given in the questionnaire did not apply to him or her. In these situations, interviewers were given the opportunity to manually enter (type) a category that the respondent felt best described him or her. The manually entered responses were called "other-specify" or "alpha-specify" responses because they were typed in a part of the question that asked the interviewer to specify an alphabetic response. These alpha-specify responses were then matched to codes to describe the responses, which were collected and maintained in a file known as a "dictionary." Other-specify responses from each survey year were matched against this file, and any responses without codes were given new codes and added to the dictionary. Consequently, the size of the dictionary file has increased each survey year.

In most cases, new unmatched responses were just new misspellings of an already established category, such as a response of "Porto Rican" instead of "Puerto Rican." If an interviewer entered both a geographical entity and a race in the other-specify response, such as "Japanese Peruvian," the geographical entity was ignored and the respondent was coded as "Japanese." The geographical entity was recorded only if no other information was available, either in the other-specify response or in the non-other-specify response. As discussed in Chapter 4, many respondents provided a race in the alpha-specify response to the Hispanic/Latino group question and vice versa, so responses to both questions were examined in the creation of each variable. This appendix summarizes the procedures that were implemented to assign race and Hispanic/Latino values to respondents based on alpha-specify responses to the questionnaire.

Once a racial category was selected that represented the other-specify response, this was combined with information that was provided in the non-other-specify categories. If the information provided in the other-specify response was so general that formal imputation seemed to be required, and more specific information was available in the non-other-specify categories, then the final assignment of a racial category was made using only the information from the non-other-specify category(ies) and the other-specify information was ignored.

C.2 Race

In the 2012 questionnaire, two core questions (QD05 and QD05ASIA) focused on the respondent's race. Respondents were permitted to select more than one race in QD05. If they selected "Asian" as one of their races, they were routed to QD05ASIA, where they were permitted to select more than one answer. Respondents had the opportunity to direct the interviewer to select "Other" as the race in both QD05 and (if applicable) QD05ASIA, whereby the interviewer then typed the alphabetic response given by the respondent. The alpha-specify responses to these two questions were considered together. The race questions used in the 2012 survey were as follows:

QD05: Which of these groups describes you?

1 White
2 Black/African American
3 American Indian/Alaska Native (American Indian includes North American, Central American, and South American Indians)
4 Native Hawaiian
5 Other Pacific Islander
6 Asian (for example: Asian Indian, Chinese, Filipino, Japanese, Korean, and Vietnamese)
7 Other (Specify)

QD05ASIA: (Asked only if level 6 of QD05 was selected.) Which of these Asian groups describes you?

1 Asian Indian
2 Chinese
3 Filipino
4 Japanese
5 Korean
6 Vietnamese
7 Other (Specify)

The Hispanic/Latino group question (QD04), discussed in Section C.3, also has an other-specify response. Whenever race information was not available from QD05 or QD05ASIA, the response to QD04 was examined to determine whether any race information was available.

C.2.1 Race Alpha Responses

The four types of race other-specify responses are described in Chapter 4. Abbreviated descriptions are repeated here for convenience.

Directly Mapped Codes. Directly mapped codes were codes mapped to one or more of the categories given in the questionnaire. There were two types of directly mapped codes: (1) racial category codes, and (2) geographic category codes. Racial category codes were exactly equivalent to one or more categories in QD05 or QD05ASIA. For example, a response such as "Han" mapped directly to a category in QD05ASIA (Chinese), and a response such as "mestizo" mapped directly to two categories in QD05 (white and American Indian/Alaska Native). Geographic category codes corresponded to a country where census data indicated a racially homogeneous society. For example, an entry of "Polish" mapped to white because the Polish census data indicated that nearly all Poles were white.

Indirectly Mapped Codes. Codes that were indirectly mapped also corresponded to countries where census data were used, but for indirect mapping, the countries were racially heterogeneous. A racial category from among the 11 categories given in the questionnaire was chosen by generating a random number and allocating the race based on a comparison of the random number with the proportions of races in the geographical entity's (country's) census. For

example, an entry of "Jamaican" would have a 76.3 percent chance of being allocated to the black/African-American category, because the latest Jamaican census indicated that 76.3 percent of Jamaicans were black. Thus, even though black Jamaicans would not consider themselves African Americans, they were allocated to the black/African-American category specified in the questionnaire. If two or three heterogeneous countries were entered in the other-specify response, the final race was allocated using the following procedure: (1) randomly assign races based on the proportions for each country mentioned, and (2) combine the results. Exceptions to these rules occurred with the categories Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain).

Starting with the 2006 imputation process, the handling of indirectly mapped codes obtained from QD05ASIA has been simplified. In earlier survey years, these types of write-in responses were mapped to a race through country census information. Since the 2006 NSDUH, however, all census-based write-in responses to the Asian race question were mapped directly to the "Other Asian" racial category.

Informative Codes for Formal Imputation Procedures. Some other-specify responses did not lead to definitive information about the respondent's race but were used to limit the final imputation. With these informative codes, the final imputation was restricted according to the information that was available. No imputation was required if more specific information was available from responses to the non-other-specify categories. For example, a response of "mixed" resulted in an imputation among donors with two or more races, and a response of "brown" resulted in an imputation among donors who were not single-race white.

Noninformative Codes. Finally, a noninformative response that was not accompanied by a response to one of the given (non-other-specify) categories resulted in an unrestricted imputation. Religious identifications (e.g., Muslim) were considered noninformative, even if the religion was usually associated with a particular ethnic group (e.g., Shinto is usually associated with Japanese).

Table C.1 lists all the race codes used in the 2012 survey, along with supplementary information related to race codes. Special situations associated with the four types of codes described in this section are discussed in the following sections. For most codes, the final assignment depended upon whether the response was given in QD05 or QD05ASIA. For informative codes described above, the six Hispanic/Latino codes—Mexican, Puerto Rican, Cuban, Central or South American, Dominican, and Spanish (from Spain)—were treated differently depending upon whether they were listed in conjunction with other racial or geographical entities.

Codes with an asterisk were those that caused the Hispanic/Latino indicator to be edited to a "yes." That is, if QD03 was either missing or "no," and any of these codes appeared as an other-specify response to QD05 or QD05ASIA, the edited Hispanic/Latino indicator (EDHOIND) was set to 1 to denote "Hispanic/Latino," and the imputation indicator for the Hispanic/Latino indicator (IIHOIND) was set to 2 to indicate "logically assigned." See Chapter 4 for more details on the edited Hispanic/Latino indicator. Note that EDHOIND also could be edited to a "no." This is discussed in Section C.3.1.

C.2.1.1 Handling of Directly Mapped Codes

For codes that were directly mapped, the final column of Table C.1 indicates to which race the code was mapped. With some exceptions, the handling of directly mapped codes that were racial categories or Asian geographic categories did not depend upon whether the response was observed in QD05 or QD05ASIA. The exceptions to this rule occurred if the response included a reference to "Indian," which was mapped to "American Indian/Alaska Native" if the response was given in QD05 and to "Asian Indian" if the response was given in QD05ASIA. On the other hand, for directly mapped codes that were non-Asian geographic categories, the final mapping always depended upon whether the response was observed in QD05 or QD05ASIA. In this case, if the code was observed in QD05ASIA, the code was always mapped to "Other Asian."

Most of the directly mapped cases were mapped directly to a single category regardless of whether the response was in QD05 or QD05ASIA. However, sometimes the category to which the code was mapped in these cases is indicated only for QD05 in the final column in Table C.1. In these instances, it was assumed that the directly mapped code for QD05ASIA was "Other Asian" (this is not shown in the table for space-saving reasons). For codes that corresponded to multiple-race respondents, individual Asian categories were not tracked.

In general, if the respondent selected one or more non-other-specify categories in QD05 and/or QD05ASIA, racial category codes were recognized, but geographic category codes were ignored. This is the primary difference in the handling of the two types of directly mapped codes. For example, if the interviewer selected the category for "black/African American" for the respondent and also wrote in "Polish," it was assumed that the respondent was a black Pole, and for racial identification purposes, the respondent was considered single-race black/African American. This was true even though the Polish census did not identify significant numbers of nonwhite persons in the Polish population.

C.2.1.2 Handling of Indirectly Mapped Codes

In most cases, indirectly mapped codes refer to heterogeneous countries where census data were used. In these cases, the race was assigned by comparing a randomly generated number to the proportion of each racial category in that country's census. As with the directly mapped codes, the final mapping of the indirectly mapped codes also depended upon whether the response was in QD05 or QD05ASIA, unless the heterogeneous countries listed were all Asian. In a similar manner to the directly mapped QD05 cases, if the code was observed in QD05ASIA, it was mapped to "Other Asian," provided none of the entries observed were Asian racial categories, Asian countries, or countries with an Asian minority. (Codes that were indirectly mapped if the response was in QD05, but were directly mapped to "Other Asian" if the response was in QD05ASIA, are denoted by "QD05ASIA: O.A." in the fourth column of the table.) Codes where there was at least one Asian minority in a specified heterogeneous country that was not all Asian, and the response was given in QD05ASIA, were handled on a case-by-case basis. The resulting strategy was either a different indirect mapping than that given if the response was in QD05 or a direct mapping.

When census data were used, it was not uncommon to find that a small proportion of the population was identified as "Other." In the rare instance that the randomly generated number indicated the respondent belonged to this "Other" group, then the selected race was determined by imputation. Codes where this was possible are identified with a superscript "I" in the third column of Table C.1. Rather than an "Other" indication, the census sometimes gave general information (e.g., Asian) where more specific information needed to be determined through imputation. In the case where the imputation was limited to Asian categories, the superscript "IA" was used.

Generally, if two entries (countries or racial categories) were observed, the race for each entity was determined first (either through a direct map or a random assignment using census data), and then the two races were combined. In some cases, a racial category was listed along with a geographical entity. As stated earlier, in most cases the geographical entity was ignored because it was usually assumed that the respondent was a resident of the listed country who also happened to be identified with the given racial category. However, it became clear on occasion that the respondent had parentage that belonged to the racial category and different parentage that came from the listed country. In these instances, the racial category was treated in the same manner as a homogeneous country of that race, and the determination of a final race was conducted in the same manner as if two countries had been listed. If one of the races listed was an Asian racial category, for example, then the response was treated in the same manner whether it was observed in QD05 or QD05ASIA. If the final assignment depended upon the census data of two indirectly mapped codes or an indirectly mapped code and a racial category, "double census" is parenthetically indicated in the third column of Table C.1. If three indirectly mapped codes were indicated by the respondent, "triple census" is indicated.[1]

Details about how to handle census information for each indirectly mapped code are shown in Table C.2. Note that the racial categories for each country listed in Table C.2 have been modified to conform to the racial categories specified by the questionnaire. For example, the black racial category from other countries has been modified to the black/African-American category. Every category and restricted imputation level with a nonzero probability of selection is listed. If a code had an indirect map (using census data) for QD05, but had a direct map for QD05ASIA, this is not specified in Table C.2. Instead, this information must be obtained from Table C.1. Explanations of the categories that are not self-explanatory are listed below.

- White or Mestizo: Imputation was restricted to respondents who were either white or Mestizo (i.e., white and American Indian/Alaska Native only). See Chapter 4 for the explanation of level 18 of EDRACE.
- Not American Indian: Imputation was restricted to respondents who were a single race other than American Indian/Alaska Native or were multiple race and American Indian/Alaska Native was not one of their component races. See Chapter 4 for the explanation of level 19 of EDRACE.

[1] When an indirectly mapped code with superscript I or IA appeared as a component in a double census or triple census code, the probability associated with the "other" category was distributed among the races appearing in the census. This was the simplest way to preserve race information from all the component indirectly mapped codes. A more complicated alternative would be to impute race information for each component country, even if the "other" category was selected at random for one or more of the race categories in the census. See the entry for "Costa Rica" (code 78) in Table C.2 for an example.

- Multiple: Imputation was restricted to respondents who were multiple race. See Chapter 4 for the explanation of level 16 of EDRACE.

C.2.1.3 Handling of Codes Informative for Formal Imputation Procedures

For six Hispanic/Latino codes that were highly prevalent in the data, census data were not used to assign the final racial category. (These are the six categories listed in QD04.) Instead, the final racial category for respondents who responded "Mexican," "Puerto Rican," "Central or South American," "Cuban," "Dominican," or "Spanish" was determined by a restricted imputation with donors who indicated one of these categories in QD04. Furthermore, if a respondent indicated any combination of these six categories, the final racial category was determined using a restricted imputation with donors who were from the geographical entities listed. On the other hand, if any of these six Hispanic/Latino groups was listed along with a second code that was not among these six, census data were used along with the census data from the second country listed. More details about how specific restricted imputations are conducted are shown in Table C.3.

If the code was observed in QD05ASIA, then the imputation was not only restricted by what was written in the other-specify response but also by the Asian categories that had the necessary attributes. Again, the information was ignored if more specific information was available from responses to the non-other-specify categories.

C.2.1.4 Noninformative Codes

For noninformative codes, a final race could still have been assigned based on responses to other categories in QD05. If no other categories were listed in QD05, race was imputed, where the imputation was restricted to a Hispanic/Latino group if the respondent gave Hispanic/Latino information in QD04. Otherwise, the final race was determined through an unrestricted imputation.

C.3 Hispanicity

As with the race questions, Hispanic/Latino respondents[2] had the opportunity to specify a Hispanic/Latino group by giving the response "Other" to QD04, the Hispanic/Latino group question. Also, respondents were permitted to select multiple Hispanic/Latino groups in response to QD04. Below is the Hispanic/Latino group question.

QD04: Which of these Hispanic, Latino, or Spanish groups best describes you?

1 Mexican/Mexican American/Mexicano/Chicano
2 Puerto Rican
3 Central or South American
4 Cuban/Cuban American
5 Dominican (from Dominican Republic)

[2] For the purposes of the instrument question-routing, Hispanic/Latino respondents were identified by their response to question QD03: "Are you of Hispanic, Latino, or Spanish origin or descent?"

6 Spanish (from Spain)
7 Other (Specify)

Levels 5 and 6 were added to QD04 after the 2004 survey. They were included because there was a large number of other-specify responses for these categories in previous years.

The QD05 and QD05ASIA questions are discussed in Section C.2. They also have other-specify responses, which were gleaned for Hispanic/Latino group information whenever no Hispanic/Latino group information was available from QD04.

C.3.1 Hispanic/Latino Group Alpha Responses

There were only two types of Hispanic/Latino group other-specify responses: (1) those that mapped to one or more EDQD04*xx*[3] variables, and (2) those that were ignored. There were no census-based routines and no responses that caused the imputation to be restricted. The imputation of a Hispanic/Latino group was restricted only when race information was available.

Table C.4 lists all the Hispanic/Latino group codes used in the 2012 survey and the Hispanic/Latino groups to which they mapped. Note that these mappings utilized the arbitrary priority rule provided in Chapter 4. This rule was used to create EDHOGRP, which skipped the intermediate step of recording the Hispanic/Latino groups indicated in QD04. These are recorded in the EDQD04*xx* variables, which are described in Chapter 4, along with the creation of EDHOGRP. The Hispanic/Latino code 600, "Stated Clearly as Not Hispanic/Latino," was unique in that it could be used to edit the Hispanic/Latino indicator, if needed. If QD03 was missing or 1, then EDHOIND was edited to a 2 if this code appeared in QD04, QD05, or QD05ASIA.

[3] See Chapter 4 for a discussion of EDQD04*xx*.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
21	White	Directly mapped (racial category)	White
22	Black/African American	Directly mapped (racial category)	Black/African American
23	American Indian/Alaska Native	Directly mapped (racial category)	American Indian/Alaska Native
24	Native Hawaiian	Directly mapped (racial category)	Native Hawaiian
25	Other Pacific Islander	Directly mapped (racial category)	Other Pacific Islander
26	Asian Indian	Directly mapped (racial category)	Asian Indian
27	Chinese	Directly mapped (racial category)	Chinese
28	Filipino	Directly mapped (racial category)	Filipino
29	Japanese	Directly mapped (racial category)	Japanese
30	Korean	Directly mapped (racial category)	Korean
31	Vietnamese	Directly mapped (racial category)	Vietnamese
32	Other Asian	Directly mapped (racial category)	Other Asian
33	Asian (nonspecific)	Codes informative for formal imputation procedures	Not a Direct Map
34	Guamanian	Directly mapped (geographic category)	Other Pacific Islander
35	Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: American Indian/Alaska Native QD05ASIA: Asian Indian
50	Belize	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.[2]
51	Guyana	QD05: Indirectly mapped[I] QD05ASIA: Directly mapped (geographic category)	QD05ASIA: Asian Indian
52	Suriname	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
53	Haiti	Indirectly mapped (QD05)	QD05ASIA: O.A.
54	Trinidad and Tobago	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
55	Jamaica	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
56	Virgin Islands (St. Thomas, St. Croix)	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
57	Bahamas	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
58	Barbados	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
59	Grenada	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
60	St. Lucia	Indirectly mapped (QD05)	QD05ASIA: O.A.
61	St. Vincent & the Grenadines	Directly mapped (geographic category)	Black/African American
62	Dominica	Directly mapped (geographic category)	Black/African American
63	Other West Indies	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
64	Brazil	Indirectly mapped (QD05)	QD05ASIA: Japanese
65	Canada	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	QD05ASIA: O.A.
66	Bahamas & Haiti	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (double census)[IA]	QD05ASIA: O.A.
67	Brazil & Portugal	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA) (double census)[IA]	QD05ASIA: O.A.
70	Mexico	Codes informative for formal imputation procedures	QD05ASIA: O.A.
71	Puerto Rico	Codes informative for formal imputation procedures	QD05ASIA: O.A.
72	Cuba	Codes informative for formal imputation procedures	QD05ASIA: O.A.
73	Dominican Republic	Codes informative for formal imputation procedures	QD05ASIA: O.A.
74	Guatemala	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
75	Honduras	Indirectly mapped (QD05)	QD05ASIA: O.A.
76	El Salvador	Indirectly mapped (QD05)	QD05ASIA: O.A.
77	Nicaragua	Indirectly mapped (QD05)	QD05ASIA: O.A.
78	Costa Rica	Indirectly mapped (QD05)[I]	QD05ASIA: Chinese
79	Panama	Indirectly mapped (QD05)	QD05ASIA: O.A.
80	Colombia	Indirectly mapped (QD05)	QD05ASIA: O.A.
81	Venezuela	Indirectly mapped (QD05)	QD05ASIA: O.A.
82	Ecuador	Indirectly mapped (QD05)	QD05ASIA: O.A.
83	Peru	Indirectly mapped (QD05)	QD05ASIA: Japanese
84	Bolivia	Indirectly mapped (QD05)	QD05ASIA: O.A.
85	Chile	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
86	Argentina	Indirectly mapped (QD05)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
87	Paraguay	Indirectly mapped (QD05)	QD05ASIA: O.A.
88	Uruguay	Indirectly mapped (QD05)	QD05ASIA: O.A.
89	Mexico & Puerto Rico	Codes informative for formal imputation procedures	Not a Direct Map
90	Mexico & Cuba	Codes informative for formal imputation procedures	Not a Direct Map
91	Mexico & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
92	Mexico & Spain	Codes informative for formal imputation procedures	Not a Direct Map
93	Puerto Rico & Cuba	Codes informative for formal imputation procedures	Not a Direct Map
94	Puerto Rico & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
95	Puerto Rico & Spain	Codes informative for formal imputation procedures	Not a Direct Map
96	Cuban & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
97	Cuban & Spain	Codes informative for formal imputation procedures	Not a Direct Map
98	Dominican & Spain	Codes informative for formal imputation procedures	Not a Direct Map
100	Norway	Directly mapped (geographic category)	QD05: White
101	Sweden	Directly mapped (geographic category)	QD05: White
102	Denmark	Directly mapped (geographic category)	QD05: White
103	United Kingdom	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
104	Ireland	Directly mapped (geographic category)	QD05: White
105	Portugal	Directly mapped (geographic category)	QD05: White
106	Spain	Codes informative for formal imputation procedures	QD05: White
107	Germany	Directly mapped (geographic category)	QD05: White
108	France	Directly mapped (geographic category)	QD05: White
109	Italy	Directly mapped (geographic category)	QD05: White
110	Netherlands	Directly mapped (geographic category)	QD05: White
111	Belgium	Directly mapped (geographic category)	QD05: White
112	Greece	Directly mapped (geographic category)	QD05: White
113	Russia	Directly mapped (geographic category)	QD05: White
114	Ukraine	Directly mapped (geographic category)	QD05: White
115	Turkey	Directly mapped (geographic category)	QD05: White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
116	Other Western Europe	Directly mapped (geographic category)	QD05: White
117	Other Eastern Europe	Directly mapped (geographic category)	QD05: White
118	Other Southern Europe	Directly mapped (geographic category)	QD05: White
119	Morocco	Directly mapped (geographic category)	QD05: White
120	Algeria	Directly mapped (geographic category)	QD05: White
121	Tunisia	Directly mapped (geographic category)	QD05: White
122	Libya	Directly mapped (geographic category)	QD05: White
123	Egypt	Directly mapped (geographic category)	QD05: White
124	Other North Africa	Directly mapped (geographic category)	QD05: White
125	Saudi Arabia	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
126	Yemen	Directly mapped (geographic category)	QD05: White
127	Oman	Directly mapped (geographic category)	QD05: White
128	UAE	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
129	Qatar	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
130	Bahrain	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
131	Israel	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
132	Iraq	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
133	Kuwait	QD05: Directly mapped (geographic category) QD05ASIA: Indirectly mapped	QD05: White
134	Iran	Directly mapped (geographic category)	Other Asian
135	Other Middle East	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
136	Armenia	Directly mapped (geographic category)	QD05: White
137	Georgia	Directly mapped (geographic category)	QD05: White
138	Azerbaijan	Directly mapped (geographic category)	QD05: White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
139	Russia Asian people groups (Tatar, Chechen, Dagestan, etc.)	Directly mapped (racial category)	Other Asian
140	Kazakhstan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
141	Uzbekistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
142	Tadjikistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
143	Kyrgizstan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
144	Turkmenistan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
145	Other Central Asia (includes Afghanistan)	Directly mapped (geographic category)	Other Asian
150	Sri Lanka	Directly mapped (geographic category)	Asian Indian
151	India	Directly mapped (geographic category)	Asian Indian
152	Other South Asia (includes Pakistan, Bangladesh, Himalayan countries)	Directly mapped (geographic category)	Asian Indian
153	Burma/Myanmar	Directly mapped (geographic category)	Other Asian
154	Laos/Hmong/Iu Mienh	Directly mapped (geographic category)	Other Asian
155	Cambodia/ Kampuchea	Directly mapped (geographic category)	Other Asian
156	Indonesia/Bali/Java	Directly mapped (geographic category)	Other Asian
157	Malaysia	Indirectly mapped[IA]	Not a Direct Map
158	Malay	Directly mapped (racial category)	QD05ASIA: O.A.
159	Singapore	Indirectly mapped[I]	Not a Direct Map
160	Thailand	Directly mapped (geographic category)	QD05ASIA: O.A.
161	Thai	Directly mapped (racial category)	QD05ASIA: O.A.
162	Mongolia	Directly mapped (geographic category)	QD05ASIA: O.A.
163	Tibet	Directly mapped (geographic category)	QD05ASIA: O.A.
164	Other East Asia	Directly mapped (geographic category)	QD05ASIA: O.A.
165	Djibouti	Indirectly mapped (QD05)	QD05ASIA: O.A.
166	Sudan	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.
167	Other Eastern Africa	Directly mapped (geographic category)	QD05ASIA: Asian Indian
168	South Africa	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
169	Namibia	Indirectly mapped (QD05)	QD05ASIA: O.A.
170	Zimbabwe	Indirectly mapped (QD05)	QD05ASIA: Asian Indian
171	Zambia	Indirectly mapped (QD05)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
172	Botswana	Directly mapped (geographic category)	QD05ASIA: O.A.
173	Angola	Indirectly mapped (QD05)	QD05ASIA: O.A.
174	Mozambique	Directly mapped (geographic category)	QD05ASIA: O.A.
175	Mauritius	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)	Not a Direct Map
176	Other Southern Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
177	Cape Verde	Indirectly mapped (QD05)	QD05ASIA: O.A.
178	Sao Tome	Directly mapped (geographic category)	QD05ASIA: O.A.
179	Mauritania	Indirectly mapped (QD05)	QD05ASIA: O.A.
180	Mali	Indirectly mapped (QD05)	QD05ASIA: O.A.
181	Niger	Indirectly mapped (QD05)	QD05ASIA: O.A.
182	Other Western Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
183	Chad	Directly mapped (geographic category)	QD05ASIA: O.A.
184	Other Central Africa	Directly mapped (geographic category)	QD05ASIA: O.A.
185	African/Africa	Directly mapped (geographic category)	QD05ASIA: Asian Indian
186	Australia	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[I]	Not a Direct Map
187	New Zealand	Indirectly mapped (specific mapping depended upon whether response was in QD05 or QD05ASIA)[IA]	Not a Direct Map
188	Fiji	Directly mapped (geographic category)	Other Pacific Islander
189	Nauru	Directly mapped (geographic category)	QD05ASIA: Chinese
190	Samoa	Indirectly mapped (QD05)	QD05ASIA: O.A.
191	Samoan	Directly mapped (racial category)	OD05ASIA: O.A.
192	Other Oceania	Directly mapped (geographic category)	QD05ASIA: O.A.
193	European (nonspecific)	Directly mapped (geographic category)	QD05ASIA: O.A.
194	Cape Verde & Portuguese	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
195	Cape Verde & Mexican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
196	Cape Verde & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
201	Biracial (nonspecific)	Codes informative for formal imputation procedures	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
202	White & Black/African American	Directly mapped (racial category)	White & Black/African American
203	White & American Indian/Alaska Native (including mestizo)	Directly mapped (racial category)	White & American Indian/Alaska Native
204	White & Native Hawaiian	Directly mapped (racial category)	White & Native Hawaiian
205	White & Other Pacific Islander	Directly mapped (racial category)	White & Other Pacific Islander
206	White & Asian Indian	Directly mapped (racial category)	White & Asian
207	White & Chinese	Directly mapped (racial category)	White & Asian
208	White & Filipino	Directly mapped (racial category)	White & Asian
209	White & Japanese	Directly mapped (racial category)	White & Asian
210	White & Korean	Directly mapped (racial category)	White & Asian
211	White & Vietnamese	Directly mapped (racial category)	White & Asian
212	White & Other Asian	Directly mapped (racial category)	White & Asian
213	White & Asian (nonspecific)	Directly mapped (racial category)	White & Asian
214	White & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White & American Indian/Alaska Native QD05ASIA: White & Asian
223	Black/African American & American Indian/Alaska Native	Directly mapped (racial category)	Black/African American & American Indian/Alaska Native
224	Black/African American & Native Hawaiian	Directly mapped (racial category)	Black/African American & Native Hawaiian
225	Black/African American & Other Pacific Islander	Directly mapped (racial category)	Black/African American & Other Pacific Islander
226	Black/African American & Asian Indian	Directly mapped (racial category)	Black/African American & Asian
227	Black/African American & Chinese	Directly mapped (racial category)	Black/African American & Asian
228	Black/African American & Filipino	Directly mapped (racial category)	Black/African American & Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
229	Black/African American & Japanese	Directly mapped (racial category)	Black/African American & Asian
230	Black/African American & Korean	Directly mapped (racial category)	Black/African American & Asian
231	Black/African American & Vietnamese	Directly mapped (racial category)	Black/African American & Asian
232	Black/African American & Other Asian	Directly mapped (racial category)	Black/African American & Asian
233	Black/African American & Asian (nonspecific)	Directly mapped (racial category)	Black/African American & Asian
234	Black/African American & Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: Black/African American & American Indian QD05ASIA: Black/African American & Asian
244	American Indian/Alaska Native & Native Hawaiian	Directly mapped (racial category)	American Indian/Alaska Native & Native Hawaiian
245	American Indian/Alaska Native & Other Pacific Islander	Directly mapped (racial category)	American Indian/Alaska Native & Other Pacific Islander
246	American Indian/Alaska Native & Asian Indian	Directly mapped (racial category)	American Indian/Alaska Native & Asian
247	American Indian/Alaska Native & Chinese	Directly mapped (racial category)	American Indian/Alaska Native & Asian
248	American Indian/Alaska Native & Filipino	Directly mapped (racial category)	American Indian/Alaska Native & Asian
249	American Indian/Alaska Native & Japanese	Directly mapped (racial category)	American Indian/Alaska Native & Asian
250	American Indian/Alaska Native & Korean	Directly mapped (racial category)	American Indian/Alaska Native & Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
251	American Indian/Alaska Native & Vietnamese	Directly mapped (racial category)	American Indian/Alaska Native & Asian
252	American Indian/Alaska Native & Other Asian	Directly mapped (racial category)	American Indian/Alaska Native & Asian
253	American Indian/Alaska Native & Asian (nonspecific)	Directly mapped (racial category)	American Indian/Alaska Native & Asian
265	Native Hawaiian & Other Pacific Islander	Directly mapped (racial category)	Native Hawaiian & Other Pacific Islander
266	Native Hawaiian & Asian Indian	Directly mapped (racial category)	Native Hawaiian & Asian
267	Native Hawaiian & Chinese	Directly mapped (racial category)	Native Hawaiian & Asian
268	Native Hawaiian & Filipino	Directly mapped (racial category)	Native Hawaiian & Asian
269	Native Hawaiian & Japanese	Directly mapped (racial category)	Native Hawaiian & Asian
270	Native Hawaiian & Korean	Directly mapped (racial category)	Native Hawaiian & Asian
271	Native Hawaiian & Vietnamese	Directly mapped (racial category)	Native Hawaiian & Asian
272	Native Hawaiian & Other Asian	Directly mapped (racial category)	Native Hawaiian & Asian
273	Native Hawaiian & Asian (nonspecific)	Directly mapped (racial category)	Native Hawaiian & Asian
286	Other Pacific Islander & Asian Indian	Directly mapped (racial category)	Other Pacific Islander & Asian
287	Other Pacific Islander & Chinese	Directly mapped (racial category)	Other Pacific Islander & Asian
288	Other Pacific Islander & Filipino	Directly mapped (racial category)	Other Pacific Islander & Asian
289	Other Pacific Islander & Japanese	Directly mapped (racial category)	Other Pacific Islander & Asian
290	Other Pacific Islander & Korean	Directly mapped (racial category)	Other Pacific Islander & Asian
291	Other Pacific Islander & Vietnamese	Directly mapped (racial category)	Other Pacific Islander & Asian
292	Other Pacific Islander & Other Asian	Directly mapped (racial category)	Other Pacific Islander & Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
293	Other Pacific Islander & Asian (nonspecific)	Directly mapped (racial category)	Other Pacific Islander & Asian
307	Asian Indian & Chinese	Directly mapped (racial category)	Multiple Asian
308	Asian Indian & Filipino	Directly mapped (racial category)	Multiple Asian
309	Asian Indian & Japanese	Directly mapped (racial category)	Multiple Asian
310	Asian Indian & Korean	Directly mapped (racial category)	Multiple Asian
311	Asian Indian & Vietnamese	Directly mapped (racial category)	Multiple Asian
312	Asian Indian & Other Asian	Directly mapped (racial category)	Multiple Asian
328	Chinese & Filipino	Directly mapped (racial category)	Multiple Asian
329	Chinese & Japanese	Directly mapped (racial category)	Multiple Asian
330	Chinese & Korean	Directly mapped (racial category)	Multiple Asian
331	Chinese & Vietnamese	Directly mapped (racial category)	Multiple Asian
332	Chinese & Other Asian	Directly mapped (racial category)	Multiple Asian
349	Filipino & Japanese	Directly mapped (racial category)	Multiple Asian
350	Filipino & Korean	Directly mapped (racial category)	Multiple Asian
351	Filipino & Vietnamese	Directly mapped (racial category)	Multiple Asian
352	Filipino & Other Asian	Directly mapped (racial category)	Multiple Asian
360	Japanese & Korean	Directly mapped (racial category)	Multiple Asian
361	Japanese & Vietnamese	Directly mapped (racial category)	Multiple Asian
362	Japanese & Other Asian	Directly mapped (racial category)	Multiple Asian
371	Korean & Vietnamese	Directly mapped (racial category)	Multiple Asian
372	Korean & Other Asian	Directly mapped (racial category)	Multiple Asian
382	Vietnamese & Other Asian	Directly mapped (racial category)	Multiple Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
383	Indian (Asian or American unclear) & Native Hawaiian	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Native Hawaiian QD05ASIA: Asian & Native Hawaiian
384	Indian (Asian or American unclear) & Other Pacific Islander	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Other Pacific Islander QD05ASIA: Asian & Other Pacific Islander
385	Indian (Asian or American unclear) & Chinese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
386	Indian (Asian or American unclear) & Filipino	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
387	Indian (Asian or American unclear) & Japanese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
388	Indian (Asian or American unclear) & Korean	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
389	Indian (Asian or American unclear) & Vietnamese	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
390	Indian (Asian or American unclear) & Other Asian	Directly mapped (racial category)	QD05: American Indian/Alaska Native & Asian QD05ASIA: Multiple Asian
401	White, Black/African American, American Indian/Alaska Native	Directly mapped (racial category)	White, Black/African American, American Indian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
402	White, Black/African American, Native Hawaiian	Directly mapped (racial category)	White, Black/African American, Native Hawaiian
403	White, Black/African American, Other Pacific Islander	Directly mapped (racial category)	White, Black/African American, Other Pacific Islander
404	White, Black/African American, Asian Indian	Directly mapped (racial category)	White, Black/African American, Asian
405	White, Black/African American, Chinese	Directly mapped (racial category)	White, Black/African American, Asian
406	White, Black/African American, Filipino	Directly mapped (racial category)	White, Black/African American, Asian
407	White, Black/African American, Japanese	Directly mapped (racial category)	White, Black/African American, Asian
408	White, Black/African American, Korean	Directly mapped (racial category)	White, Black/African American, Asian
409	White, Black/African American, Vietnamese	Directly mapped (racial category)	White, Black/African American, Asian
410	White, Black/African American, Other Asian	Directly mapped (racial category)	White, Black/African American, Asian
411	White, Black/African American, Asian (nonspecific)	Directly mapped (racial category)	White, Black/African American, Asian
412	White, American Indian/Alaska Native, Native Hawaiian	Directly mapped (racial category)	White, American Indian/Alaska Native, Native Hawaiian
413	White, American Indian/Alaska Native, Other Pacific Islander	Directly mapped (racial category)	White, American Indian/Alaska Native, Other Pacific Islander
414	White, American Indian/Alaska Native, Asian Indian	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
415	White, American Indian/Alaska Native, Chinese	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
416	White, American Indian/Alaska Native, Filipino	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
417	White, American Indian/Alaska Native, Japanese	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
418	White, American Indian/Alaska Native, Korean	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
419	White, American Indian/Alaska Native, Vietnamese	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
420	White, American Indian/Alaska Native, Other Asian	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
421	White, American Indian/Alaska Native, Asian (nonspecific)	Directly mapped (racial category)	White, American Indian/Alaska Native, Asian
422	White, Black/African American, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Black/African American, American Indian/Alaska Native QD05ASIA: White, Black/African American, Asian
423	White, Native Hawaiian, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Native Hawaiian, American Indian/Alaska Native QD05ASIA: White, Native Hawaiian, Asian
424	White, Other Pacific Islander, Indian (Asian or American unclear)	Directly mapped (racial category)	QD05: White, Other Pacific Islander, American Indian/Alaska Native QD05ASIA: White, Other Pacific Islander, Asian
600	Mexican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
601	Mexican & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
602	Mexican & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
603	Mexican & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
604	Mexican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
605	Mexican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
606	Mexican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
607	Mexican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
608	Mexican & Ecuadorian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
609	Mexican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
610	Mexican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
611	Mexican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
612	Mexican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
613	Mexican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
614	Mexican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
615	Mexican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
616	Puerto Rican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
617	Puerto Rican & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
618	Puerto Rican & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
619	Puerto Rican & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
620	Puerto Rican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
621	Puerto Rican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
622	Puerto Rican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
623	Puerto Rican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
624	Puerto Rican & Ecuadorian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
625	Puerto Rican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
626	Puerto Rican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
627	Puerto Rican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
628	Puerto Rican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
629	Puerto Rican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
630	Puerto Rican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
631	Puerto Rican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
632	Cuban & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
633	Cuban & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
634	Cuban & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
635	Cuban & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
636	Cuban & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
637	Cuban & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
638	Cuban & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
639	Cuban & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
640	Cuban & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
641	Cuban & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
642	Cuban & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
643	Cuban & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
644	Cuban & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
645	Cuban & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
646	Cuban & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
647	Cuban & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
648	Dominican & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
649	Dominican & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
650	Dominican & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
651	Dominican & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
652	Dominican & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
653	Dominican & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
654	Dominican & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
655	Dominican & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
656	Dominican & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
657	Dominican & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
658	Dominican & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
659	Dominican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
660	Dominican & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
661	Dominican & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
662	Dominican & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
663	Dominican & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
664	Spanish (from Spain) & Guatemalan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
665	Spanish (from Spain) & El Salvadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
666	Spanish (from Spain) & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
667	Spanish (from Spain) & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
668	Spanish (from Spain) & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
669	Spanish (from Spain) & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
670	Spanish (from Spain) & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
671	Spanish (from Spain) & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
672	Spanish (from Spain) & Ecuadorian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
673	Spanish (from Spain) & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
674	Spanish (from Spain) & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
675	Spanish (from Spain) & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
676	Spanish (from Spain) & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
677	Spanish (from Spain) & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
678	Spanish (from Spain) & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
679	Spanish (from Spain) & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
680	Colombian & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
681	Colombian & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
682	Colombian & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
683	Colombian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
684	Colombian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
685	Colombian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
686	Colombian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
687	Colombian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
688	Colombian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
689	Venezuelan & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
690	Venezuelan & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
691	Venezuelan & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
692	Venezuelan & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
693	Venezuelan & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
694	Venezuelan & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
695	Venezuelan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
696	Venezuelan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
697	Ecuadoran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
698	Ecuadoran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
699	Ecuadoran & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
700	Ecuadoran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
701	Ecuadoran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
702	Ecuadoran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
703	Ecuadoran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
704	Peruvian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
705	Peruvian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
706	Peruvian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
707	Peruvian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
708	Peruvian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
709	Peruvian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
710	Bolivian & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
711	Bolivian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
712	Bolivian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
713	Bolivian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
714	Bolivian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
715	Chilean & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
716	Chilean & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
717	Chilean & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
718	Chilean & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
719	Argentine & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
720	Argentine & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
721	Argentine & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
722	Paraguayan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
723	Paraguayan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
724	Uruguayan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
725	Guatemalan & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
726	El Salvadoran & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
727	Honduran & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
728	Nicaraguan & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
729	Costa Rican & Colombian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
730	Panamanian & Colombian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
731	Guatemalan & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
732	El Salvadoran & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
733	Honduran & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
734	Nicaraguan & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
735	Costa Rican & Venezuelan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
736	Panamanian & Venezuelan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
737	Guatemalan & Ecuadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
738	El Salvadoran & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
739	Honduran & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
740	Nicaraguan & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
741	Costa Rican & Ecuadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
742	Panamanian & Ecuadoran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
743	Guatemalan & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
744	El Salvadoran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
745	Honduran & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
746	Nicaraguan & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
747	Costa Rican & Peruvian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
748	Panamanian & Peruvian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
749	Guatemalan & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
750	El Salvadoran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
751	Honduran & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
752	Nicaraguan & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
753	Costa Rican & Bolivian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
754	Panamanian & Bolivian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
755	Guatemalan & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
756	El Salvadoran & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
757	Honduran & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
758	Nicaraguan & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
759	Costa Rican & Chilean	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
760	Panamanian & Chilean	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
761	Guatemalan & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
762	El Salvadoran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
763	Honduran & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
764	Nicaraguan & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
765	Costa Rican & Argentine	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
766	Panamanian & Argentine	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
767	Guatemalan & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
768	El Salvadoran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
769	Honduran & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
770	Nicaraguan & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
771	Costa Rican & Paraguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
772	Panamanian & Paraguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
773	Guatemalan & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
774	El Salvadoran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
775	Honduran & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
776	Nicaraguan & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
777	Costa Rican & Uruguayan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
778	Panamanian & Uruguayan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
779	Guatemalan & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
780	El Salvadoran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
781	Honduran & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
782	Nicaraguan & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
783	Costa Rican & Brazilian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
784	Panamanian & Brazilian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
785	Guatemalan & El Salvadoran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
786	Guatemalan & Honduran	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
787	Guatemalan & Nicaraguan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
788	Guatemalan & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
789	Guatemalan & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
790	El Salvadoran & Honduran	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
791	El Salvadoran & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
792	El Salvadoran & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
793	El Salvadoran & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
794	Honduran & Nicaraguan	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
795	Honduran & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
796	Honduran & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
797	Nicaraguan & Costa Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
798	Nicaraguan & Panamanian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
799	Costa Rican & Panamanian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
800	Mexican & Jamaican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
801	Puerto Rican & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
802	Cuban & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
803	Dominican & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
804	Spanish (from Spain) & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
805	Mexican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
806	Puerto Rican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
807	Cuban & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
808	Dominican & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
809	Spanish (from Spain) & Other European	Directly mapped (geographic category)	QD05: White
810	Trinidadian & Mexican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
811	Trinidadian & Puerto Rican	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
812	Trinidadian & Cuban	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
813	Trinidadian & Dominican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
814	Trinidadian & Spanish (from Spain)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
815	Guatemalan & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
816	El Salvador & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
817	Honduran & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
818	Nicaraguan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
819	Costa Rican & European (not Spanish)	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
820	Panamanian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
821	Colombian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
822	Venezuelan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
823	Ecuadoran & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
824	Peruvian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
825	Bolivian & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
826	Chilean & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
827	Argentine & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
828	Paraguay & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
829	Uruguayan & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
830	Brazil & European (not Spanish)	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
831	(part) Mexican, ½ (part) White	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
832	(part) Mexican, ½ (part) Black/African American	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
833	(part) Mexican, ½ (part) American Indian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
834	(part) Mexican, ½ (part) Hawaiian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
835	(part) Mexican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
836	(part) Mexican, ½ (part) Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
837	(part) Mexican, ½ (part) Chinese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
838	(part) Mexican, ½ (part) Filipino	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
839	(part) Mexican, ½ (part) Japanese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
840	(part) Mexican, ½ (part) Korean	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
841	(part) Mexican, ½ (part) Vietnamese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
842	(part) Mexican, ½ (part) Other Asian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
843	(part) Puerto Rican, ½ (part) White	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
844	(part) Puerto Rican, ½ (part) Black/African American	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
845	(part) Puerto Rican, ½ (part) American Indian/Alaska Native	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
846	(part) Puerto Rican, ½ (part) Native Hawaiian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
847	(part) Puerto Rican, ½ (part) Other Pacific Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
848	(part) Puerto Rican, ½ (part) Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
849	(part) Puerto Rican, ½ (part) Chinese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
850	(part) Puerto Rican, ½ (part) Filipino	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
851	(part) Puerto Rican, ½ (part) Japanese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
852	(part) Puerto Rican, ½ (part) Korean	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
853	(part) Puerto Rican, ½ (part) Vietnamese	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
854	(part) Puerto Rican, ½ (part) Other Asian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
855*	(part) Hispanic/Latino, ½ (part) White	Directly mapped (racial category)	White

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
856*	(part) Hispanic/Latino, ½ (part) Black/African American	Directly mapped (racial category)	Black/African American
857*	(part) Hispanic/Latino, ½ (part) American Indian/Alaska Native	Directly mapped (racial category)	American Indian/Alaska Native
858*	(part) Hispanic/Latino, ½ (part) Native Hawaiian	Directly mapped (racial category)	Native Hawaiian
859*	(part) Hispanic/Latino, ½ (part) Other Pacific Islander	Directly mapped (racial category)	Other Pacific Islander
860*	(part) Hispanic/Latino, ½ (part) Asian Indian	Directly mapped (racial category)	Asian Indian
861*	(part) Hispanic/Latino, ½ (part) Chinese	Directly mapped (racial category)	Chinese
862*	(part) Hispanic/Latino, ½ (part) Filipino	Directly mapped (racial category)	Filipino
863*	(part) Hispanic/Latino, ½ (part) Japanese	Directly mapped (racial category)	Japanese
864*	(part) Hispanic/Latino, ½ (part) Korean	Directly mapped (racial category)	Korean
865*	(part) Hispanic/Latino, ½ (part) Vietnamese	Directly mapped (racial category)	Vietnamese
866*	(part) Hispanic/Latino, ½ (part) Other Asian	Directly mapped (racial category)	Other Asian
869	Haitian & Jamaican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
870	Haitian & Dominican	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
871	Honduran & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
872	Guatemalan & Iranian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
873	Panamanian & Jamaican	Indirectly mapped (QD05) (double census)I	QD05ASIA: O.A.
874	Cuban & Thai	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
875	Venezuelan & Trinidadian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
876	Puerto Rican & Arab	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
877	Puerto Rican & Virgin Islander	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
878	Mexican & Samoan	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
879	Salvadoran & Egyptian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
880	Costa Rican & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
881	Mexican & Iranian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
882	Spanish & Barbadian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
883	Peruvian & Other Middle Eastern	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
884	Puerto Rican & African	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
885	Jamaican & Egyptian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
886	Argentine & Turkish	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
887	Mexican & Egyptian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
888	Guatemalan & Canadian	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
889	Haitian & European	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
890	Argentine, Cuban, & Spanish	Indirectly mapped (QD05) (triple census)	QD05ASIA: O.A.
891	Mexican, Cuban, & France	Indirectly mapped (QD05) (triple census)[I]	QD05ASIA: O.A.
892	Mexican, Puerto Rican, & European	Indirectly mapped (QD05) (triple census)[I]	QD05ASIA: O.A.
893	Haiti & Trinidad	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
894	Belize & Honduras	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
895	Trinidad & European	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.
896	Puerto Rican & Haitian	Indirectly mapped (QD05) (double census)	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
900*	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain or Dominican Republic)	Codes informative for formal imputation procedures	Not a Direct Map
901*	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Codes informative for formal imputation procedures	Not a Direct Map
902*	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	Indirectly mapped (QD05)	QD05ASIA: Other Asian
903	Central/South American (no country)	Codes informative for formal imputation procedures	Not a Direct Map
904	Nonwhite (nonspecific/brown)	Codes informative for formal imputation procedures	Not a Direct Map
905*	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Codes informative for formal imputation procedures	Not a Direct Map
906*	Mezclado, Mezclada (Hispanic/Latino mixed)	Codes informative for formal imputation procedures	Not a Direct Map
907	Mixed	Codes informative for formal imputation procedures	Not a Direct Map
908	Olive	Directly mapped (geographic category)	White
909	Creole	Indirectly mapped	QD05ASIA: O.A.
910	Arab	Directly mapped (geographic category)	QD05: White QD05ASIA: Other Asian
911	Jewish	Directly mapped (geographic category)	White
912	Kurd	Directly mapped (geographic category)	Other Asian
913	Chaldean/Caldanian/ Assyrian	Directly mapped (geographic category)	Other Asian
914	Romany/Gypsy	Directly mapped (geographic category)	White
915	Central/South American & West Indies	Indirectly mapped	QD05ASIA: O.A.

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
916	Central/South American & Mexican	Codes informative for formal imputation procedures	Not a Direct Map
917	Central/South American & Puerto Rican	Codes informative for formal imputation procedures	Not a Direct Map
918	Central/South American & Cuban	Codes informative for formal imputation procedures	Not a Direct Map
919	Central/South American & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
920	Central/South American & Spanish	Codes informative for formal imputation procedures	Not a Direct Map
921	Arab/Asian	QD05: Directly mapped (racial category) QD05ASIA: Directly mapped (geographic category)	QD05: White & Asian QD05ASIA: Other Asian
922	Arab/European	Directly mapped (geographic category)	QD05: White QD05ASIA: White & Asian
923	Arab/African	Directly mapped (geographic category)	QD05: White & Black/African American QD05ASIA: Asian & Black/African American
924	Arab/Chaldean	Directly mapped (racial category)	QD05: White & Asian QD05ASIA: Other Asian
925	European & Asian Indian	Directly mapped (racial category)	White & Asian
926	West Indies & Belize	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
927	West Indies & Cape Verde	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
928	Arab & Cape Verde	Indirectly mapped (QD05) (double census)[I]	QD05ASIA: O.A.
929	Aryan	Directly mapped (geographic category)	QD05: White QD05ASIA: Asian Indian
930	Turkish & Lebanese	Directly mapped (racial category)	White
932	Puerto Rican & Dominican	Codes informative for formal imputation procedures	Not a Direct Map
933	Puerto Rican & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
934	Dominican & Asian Indian	Indirectly mapped (QD05) (double census)[I]	Not a Direct Map
935	Cuban & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
936	Dominican & Spanish (from Spain)	Codes informative for formal imputation procedures	Not a Direct Map
951	White and something else	Directly mapped (racial category)	White (Multiple Race)
952	Black/African American and something else	Directly mapped (racial category)	Black/African American (Multiple Race)
953	American Indian/Alaska Native and something else	Directly mapped (racial category)	American Indian/Alaska Native (Multiple Race)
954	Native Hawaiian and something else	Directly mapped (racial category)	Native Hawaiian (Multiple Race)
955	Other Pacific Islander and something else	Directly mapped (racial category)	Other Pacific Islander (Multiple Race)
956	Asian Indian and something else	Directly mapped (racial category)	Asian Indian (Multiple Race)
957	Chinese and something else	Directly mapped (racial category)	Chinese (Multiple Race)
958	Filipino and something else	Directly mapped (racial category)	Filipino (Multiple Race)
959	Japanese and something else	Directly mapped (racial category)	Japanese (Multiple Race)
960	Korean and something else	Directly mapped (racial category)	Korean (Multiple Race)
961	Vietnamese and something else	Directly mapped (racial category)	Vietnamese (Multiple Race)
962	Other Asian and something else	Directly mapped (racial category)	Other Asian (Multiple Race)
963	Asian (nonspecific) and something else	Codes useful for formal imputation procedures	Not a Direct Map
964	Indian (Asian or American unclear) and something else	Directly mapped (racial category)	QD05: American Indian/Alaska Native (Multiple Race) QD05ASIA: Asian Indian (Multiple Race)
965	Brown & White	Directly mapped (racial category)	White & Black/African American
985	Bad data	Noninformative code	Not a Direct Map
994	Unknown/"Don't Know"	Noninformative code	Not a Direct Map

Table C.1 Descriptions of Race Codes and the Categories to Which They Mapped (continued)

Race Code	Race Name	Type[1]	Category to Which Race Code Directly Mapped
997	Rather Not Say/"Refused" ("American" or "All of Them")	Noninformative code	Not a Direct Map

* These codes caused the Hispanic/Latino indicator to be edited to a "yes" if QD03 was missing or "no." The code that caused the Hispanic/Latino indicator to be edited to a "no" was a Hispanic/Latino code (600) and is listed in Table C.4.

[1] Among the indirectly mapped codes, codes where an imputation was possible based on census information are indicated by the superscript "I." If the imputation was limited to Asians in these cases, the superscript "IA" is used. See Section C.2.1.2 for details.

[2] The abbreviation "O.A." is equivalent to "Other Asian."

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes

Race Code	Race Name	Probabilities
50	Belize	6.1% Black/African American, 10.6% American Indian/Alaska Native, 24.9% White and Black/African American, 48.7% White and American Indian/Alaska Native, 9.7% Unrestricted Imputation
51	Guyana	QD05: 36% Black/African American, 7% American Indian/Alaska Native, 50% Asian Indian, 7% Unrestricted Imputation
52	Suriname	QD05: 1% White, 10% Black/African American, 2% American Indian/Alaska Native, 37% Asian Indian, 2% Chinese, 15% Other Asian, 31% White and Black/African American, 2% Unrestricted Imputation QD05ASIA: 71% Asian Indian, 29% Other Asian
53	Haiti	95% Black/African American, 5% White and Black/African American
54	Trinidad and Tobago	QD05: 0.6% White, 39.5% Black/African American, 40.3% Asian Indian, 1.2% Chinese, 18.4% Black/African American and Asian Indian QD05ASIA: 69% Asian Indian, 31% Black/African American and Asian Indian
55	Jamaica	QD05: 3.2% White, 76.3% Black/African American, 1.5% Asian Indian, .6% Chinese, 15.1% White and Black/African American, 1.5% Black/African American and Asian Indian, .6% Black/African American and Chinese, 1.2% Unrestricted Imputation QD05ASIA: 36% Asian Indian, 36% Black/African American and Asian Indian, 14% Chinese, 14% Black/African American and Chinese
56	Virgin Is (St Thomas, St Croix)	QD05: 12% White, 85% Black/African American, 3% Asian Nonspecific QD05ASIA: Impute among Asians
57	Bahamas	QD05: 12% White, 85% Black/African American, 3% Asian Nonspecific QD05ASIA: Impute among Asians
58	Barbados	4% White, 90% Black/African American, 6% Unrestricted Imputation
59	Grenada	82% Black/African American, 13% White and Black/African American, 5% Unrestricted Imputation
60	St. Lucia	QD05: 1% White, 90% Black/African American, 3% Asian Indian, 3% White and Black/African American, 3% Black/African American and Asian Indian QD05ASIA: 50% Asian Indian 50% Black/African American and Asian Indian
63	Other West Indies	QD05: 80% Black/African American, 14% Asian Nonspecific, 6% Unrestricted Imputation QD05ASIA: Impute among Asians
64	Brazil	55.3% White, 6% Black/African American, .3% American Indian/Alaska Native, .3% Japanese, 38% White and Black/African American
65	Canada	QD05: 66% White, 2% American Indian/Alaska Native, 32% Unrestricted Imputation QD05ASIA: Impute among Asians

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
70	Mexico	9.3% White, 30.3% American Indian/Alaska Native, 60.3% White and American Indian/Alaska Native
71	Puerto Rico	QD05: 82.7% White, 10.2% Black/African American, .4% American Indian/Alaska Native, .01% Native Hawaiian, .02% Other Pacific Islander, .13% Asian Indian, .05% Chinese, .01% Filipino, .01% Japanese, .01% Korean, .01% Vietnamese, 6.4% White and Black/African American QD05ASIA: 59% Asian Indian, 23% Chinese, 4.5% each Filipino, Japanese, Korean, Vietnamese
72	Cuba	37% White, 11% Black/African American, 1% Chinese, 51% White and Black/African American
73	Dominican Republic	16% White, 11% Black/African American, 73% White and Black/African American
74	Guatemala	43% American Indian/Alaska Native, 55% White and American Indian/Alaska Native, 2% Unrestricted Imputation
75	Honduras	1% White, 2% Black/African American, 7% American Indian/Alaska Native, 90% White and American Indian/Alaska Native
76	El Salvador	9% White, 1% American Indian/Alaska Native, 90% White and American Indian/Alaska Native
77	Nicaragua	17% White, 9% Black/African American, 5% American Indian/Alaska Native, 69% White and American Indian/Alaska Native
78	Costa Rica	QD05: 3% Black/African American, 1% American Indian/Alaska Native, 1% Chinese, 94% White or Mestizo, 1% Unrestricted Imputation QD05 when in combination with another race: 47.2% White, 3.2% Black/African American, 1.2% American Indian/Alaska Native, 1.2% Chinese, 47.2% White and American Indian/Alaska Native
79	Panama	10% White, 14% Black/African American, 6% American Indian/Alaska Native, 70% White and American Indian/Alaska Native
80	Colombia	20% White, 4% Black/African American, 1% American Indian/Alaska Native, 14% White and Black/African American, 58% White and American Indian/Alaska Native, 3% Black/African American and American Indian/Alaska Native
81	Venezuela	21% White, 10% Black/African American, 2% American Indian/Alaska Native, 67% White and American Indian/Alaska Native
82	Ecuador	7% White, 3% Black/African American, 25% American Indian/Alaska Native, 65% White and American Indian/Alaska Native
83	Peru	15% White, 1% Black/African American, 45% American Indian/Alaska Native, 1% Chinese, 1% Japanese, 37% White and American Indian/Alaska Native
84	Bolivia	15% White, 55% American Indian/Alaska Native, 30% White and American Indian/Alaska Native

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
85	Chile	3% American Indian/Alaska Native, 95% White or Mestizo, 2% Unrestricted Imputation
86	Argentina	97% White, 3% White and American Indian/Alaska Native
87	Paraguay	2.5% White, 2.5% American Indian/Alaska Native, 95% White and American Indian/Alaska Native
88	Uruguay	88% White, 4% Black/African American, 8% White and American Indian/Alaska Native
103	United Kingdom	97.2% White, 1.4% Black/African American, 1.4% Asian Indian
125	Saudi Arabia	QD05: 90% White, 10% Asian Indian QD05ASIA: 90% Other Asian, 10% Asian Indian
128	UAE	QD05: 30.5% White, 50% Asian Indian, 11.5% Other Asian, 8% Not American Indian/Alaska Native QD05ASIA: 50% Asian Indian, 50% Other Asian
129	Qatar	QD05: 40% White, 36% Asian Indian, 10% Other Asian, 14% Not American Indian/Alaska Native QD05ASIA: 36% Asian Indian, 64% Other Asian
130	Bahrain	QD05: 73% White, 8% Other Asian, 19% Asian Nonspecific QD05ASIA: 81% Other Asian, 19% Impute among Asian Groups
133	Kuwait	QD05ASIA: 9% Asian Indian, 91% Other Asian
140	Kazakhstan	36.1% White, 57.3% Other Asian, 6.6% Not American Indian/Alaska Native
141	Uzbekistan	5.5% White, 92% Other Asian, 2.5% Not American Indian/Alaska Native
142	Tadjikistan	3.5% White, 89.9% Other Asian, 6.6% Not American Indian/Alaska Native
143	Kyrgizstan	22.9% White, 65.3% Other Asian, 11.8% Not American Indian/Alaska Native
144	Turkmenistan	6.7% White, 88.2% Other Asian, 5.1% Not American Indian/Alaska Native
157	Malaysia	8% Asian Indian, 24% Chinese, 58% Other Asian, 10% Asian Nonspecific
159	Singapore	7.9% Asian Indian, 76.7% Chinese, 14% Other Asian, 1.4% Not American Indian/Alaska Native
165	Djibouti	2.5% White, 97.5% Black/African American
166	Sudan	39% White, 58% Black/African American, 3% Not American Indian/Alaska Native
168	South Africa	13.6% White, 75.2% Black/African American, 2.6% Asian Indian, 8.6% White and Black/African American
169	Namibia	6% White, 87.5% Black/African American, 6.5% White and Black/African American
170	Zimbabwe	1% White, 98% Black/African American, .5% Asian Indian, .5% White and Black/African American

Table C.2 Proportional Racial Allocations for Indirectly Mapped Codes (continued)

Race Code	Race Name	Probabilities
171	Zambia	1.1% White, 98.7% Black/African American, .2% Not American Indian/Alaska Native
173	Angola	1% White, 97% Black/African American, 2% White and Black/African American
175	Mauritius	QD05: 2% White, 68% Asian Indian, 3% Chinese, 27% White and Black/African American QD05ASIA: 96% Asian Indian, 4% Chinese
177	Cape Verde	1% White, 28% Black/African American, 71% White and Black/African American
179	Mauritania	30% White, 30% Black/African American, 40% White and Black/African American
180	Mali	10% White, 90% Black/African American
181	Niger	9% White, 91% Black/African American
186	Australia	QD05: 92% White, 7% Asian Nonspecific, 1% Not American Indian/Alaska Native QD05ASIA: Impute among Asians
187	New Zealand	QD05: 79.1% White, 13.5% Other Pacific Islander, 7.4% Asian Nonspecific QD05ASIA: Impute among Asians
190	Samoa	.4% White, 92.6% Other Pacific Islander, 7% White and Other Pacific Islander
902	Definitely Hispanic/Latino (Hispanic Dominican Republic, Dominicano, etc.)	16% White, 11% Black/African American, 73% White and Black/African American
909	Creole	50% White, 50% White and Black/African American
915	Central/South American & West Indies	50% White and Black/African American, 50% Black/African American and American Indian/Alaska Native
920	Central/South American & Spanish	50% White, 50% White and American Indian/Alaska Native

Table C.3 Procedures for Restricted Imputation for Codes Informative for Formal Imputation Procedures

Race Code	Race Name	Restriction on Donors in Formal Imputation
33	Asian (nonspecific)	Donors were Asian: impute specific Asian group
70	Mexico	Donors were Mexican[1]
71	Puerto Rico	Donors were Puerto Rican
72	Cuba	Donors were Cuban
78	Costa Rica (QD05: 94% White or Mestizo)	For this 94%, donors were white or white and American Indian/Alaska Native
89	Mexico & Puerto Rico	Donors were Mexican, Puerto Rican, or both
90	Mexico & Cuba	Donors were Mexican, Cuban, or both
91	Mexico & Dominican Republic	Donors were Mexican, Dominican, or both
92	Mexico & Spain	Donors were Mexican, Spanish, or both
93	Puerto Rico & Cuba	Donors were Puerto Rican, Cuban, or both
94	Puerto Rico & Dominican Republic	Donors were Puerto Rican, Dominican, or both
95	Puerto Rico & Spain	Donors were Puerto Rican, Spanish, or both
96	Cuba & Dominican Republic	Donors were Cuban, Dominican, or both
97	Cuba & Spain	Donors were Cuban, Spanish, or both
98	Dominican Republic & Spain	Donors were Dominican, Spanish, or both
128	UAE (QD05: 8% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
129	Qatar (QD05: 14% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
140	Kazakhstan (QD05: 6.6% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
141	Uzbekistan (QD05: 2.5% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
142	Tadjikistan (QD05: 6.6% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
143	Kyrgizstan (QD05: 11.8% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
144	Turkmenistan (QD05: 5.1% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
159	Singapore (QD05: 1.4% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
166	Sudan (QD05: 3% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
171	Zambia (QD05: 0.2% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
186	Australia (QD05: 1% Not American Indian/Alaska Native)	Donors included respondents of any race or races that did not include American Indian/Alaska Native
201	Biracial (nonspecific)	Donors were multiple race: imputed constituent races[2]

Table C.3 Procedures for Restricted Imputation for Codes Informative for Formal Imputation Procedures (continued)

Race Code	Race Name	Restriction on Donors in Formal Imputation
900	Definitely Hispanic/Latino (Hispanic, Latino/a, Chicano/a, etc., not Spain, D.R.)	Donors were Hispanic/Latino
901	Definitely Hispanic/Latino (Hispanic Spanish, Español, etc.)	Donors were Hispanic/Latino
903	Central/South American (no country)	Donors were Central/South American
904	Nonwhite (nonspecific/brown)	Donors were any race but single-race white
905	Hispanic/Latino nonwhite (including trigueno = "dark," moreno)	Donors were Hispanic/Latino who were any race but single-race white
906	Mezclado, Mezclada (Hispanic/Latino mixed)	Donors were multiple race and Hispanic/Latino: imputed constituent races
907	Mixed	Donors were multiple race: imputed constituent races
916	Central/South American & Mexican	Donors were Central/South American, Mexican, or both
917	Central/South American & Puerto Rican	Donors were Central/South American, Puerto Rican, or both
918	Central/South American & Cuban	Donors were Central/South American, Cuban, or both
919	Central/South American & Dominican	Donors were Central/South American, Dominican, or both
920	Central/South American & Spanish	Donors were Central/South American, Spanish, or both
932	Puerto Rican & Dominican	Donors were Puerto Rican, Dominican, or both
933	Puerto Rican & Spanish (from Spain)	Donors were Puerto Rican, Spanish, or both
935	Cuban & Spanish (from Spain)	Donors were Cuban, Spanish, or both
936	Dominican & Spanish (from Spain)	Donors were Dominican, Spanish, or both

[1] Even though a recipient may not be Hispanic/Latino, he or she may still have indicated "Mexican" in the QD05 other-specify response. Donors in this case included both Hispanic/Latino and (though extremely rare) non-Hispanic/Latino Mexicans.

[2] Because most multiple-race respondents have only two constituent races, any respondent with this code and nothing else is likely to be assigned a biracial donor. However, for the sake of simplicity, respondents with this code were not treated any differently than respondents with code 907 ("Mixed").

Table C.4 Mapping of Hispanic/Latino Group Codes

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
11	Mexican/Mexican American/Mexicano/Chicano	Mexican
12	Puerto Rican	Puerto Rican
13	Central or South American	Central or South American
14	Cuban/Cuban American	Cuban
15	Dominican (Dominican Republic)	Dominican
16	Spanish (from Spain)	Spanish
17	Caribbean Hispanic/Latino (not specified as Dominican)	Other Hispanic
21	Mexican & Puerto Rican	Mexican
22	Mexican & Central or South American	Mexican
23	Mexican & Cuban	Mexican
24	Mexican & Dominican	Mexican
25	Mexican & Spanish (from Spain)	Mexican
26	Puerto Rican & Central or South American	Puerto Rican
27	Puerto Rican & Cuban	Cuban
28	Puerto Rican & Dominican	Puerto Rican
29	Puerto Rican & Spanish (from Spain)	Puerto Rican
30	Central or South American & Cuban	Cuban
31	Central or South American & Dominican	Central or South American
32	Central or South American & Spanish (from Spain)	Central or South American
33	Cuban & Dominican	Cuban
34	Cuban & Spanish (from Spain)	Cuban
35	Dominican & Spanish (from Spain)	Dominican
36	Mexican, Puerto Rican, & Central or South American	Mexican
37	Mexican, Puerto Rican, & Cuban	Mexican
38	Mexican, Puerto Rican, & Dominican	Mexican
39	Mexican, Puerto Rican, & Spanish (from Spain)	Mexican
40	Mexican, Central or South American, & Cuban	Mexican
41	Mexican, Central or South American, & Dominican	Mexican
42	Mexican, Central or South American, & Spanish (from Spain)	Mexican
43	Mexican, Cuban, & Dominican	Mexican
44	Mexican, Cuban, & Spanish (from Spain)	Mexican
45	Mexican, Dominican, & Spanish (from Spain)	Mexican
46	Puerto Rican, Central or South American, & Cuban	Cuban
47	Puerto Rican, Central or South American, & Dominican	Puerto Rican

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
48	Puerto Rican, Central or South American, & Spanish (from Spain)	Puerto Rican
49	Puerto Rican, Cuban, & Dominican	Cuban
50	Puerto Rican, Cuban, & Spanish (from Spain)	Cuban
51	Puerto Rican, Dominican, & Spanish (from Spain)	Puerto Rican
52	Central or South American, Cuban, & Dominican	Cuban
53	Central or South American, Cuban, & Spanish (from Spain)	Cuban
54	Central or South American, Dominican, & Spanish (from Spain)	Central or South American
55	Cuban, Dominican, and Spanish (from Spain)	Cuban
56	Portuguese & Mexican	Mexican
57	Portuguese & Puerto Rican	Puerto Rican
58	Portuguese & Cuban	Cuban
59	Portuguese & Central or South American	Central or South American
60	Portuguese & Dominican	Dominican
61	Portuguese & Spanish (from Spain)	Spanish
100	Brazilian	Central or South American
101	Portuguese	Other Hispanic/Latino
102	Cape Verde	Other Hispanic/Latino
103	Belizean (formerly British Honduras)	Central or South American
104	Guyana	Central or South American
105	Jamaican	Other Hispanic/Latino
106	Other Caribbean (possibly Hispanic)	Other Hispanic/Latino
107	Philippines & Guam	Other Hispanic/Latino
108	Brazilian & Portuguese	Central or South American
109	Cape Verde & Portuguese	Other Hispanic/Latino
110	Haitian/Jamaican	Other Hispanic/Latino
200	Mexican/Jamaican	Mexican
201	Puerto Rican/Jamaican	Puerto Rican
202	Central or South American/Jamaican	Central or South American
203	Cuban/Jamaican	Cuban
204	Dominican/Jamaican	Dominican
205	Spanish (from Spain)/Jamaican	Spanish
206	Mexican/West Indies	Mexican
207	Puerto Rican/West Indies	Puerto Rican
208	Central or South American/West Indies	Central or South American

Table C.4 Mapping of Hispanic/Latino Group Codes (continued)

Hispanic/Latino Code	Hispanic/Latino Group Name	Category to Which Hispanic/Latino Code Directly Mapped
209	Cuban/West Indies	Cuban
210	Dominican/West Indies	Dominican
211	Spanish (from Spain)/West Indies	Spanish
212	Mexican/Haitian	Mexican
213	Puerto Rican/Haitian	Puerto Rican
214	Central or South American/Haitian	Central or South American
215	Cuban/Haitian	Cuban
216	Dominican/Haitian	Dominican
217	Spanish (from Spain)/Haitian	Spanish
500	Hispanic/Latino	Hispanic/Latino group imputed
501	Hispanic/Latino Mixed/Mezclada	Hispanic group imputed
502	Hispanic Creole	Other Hispanic
600*	Stated clearly as Not Hispanic/Latino	Hispanic/Latino indicator edited to "no"
800	Non-Hispanic/Latino country	Other Hispanic/Latino
801	Racial category (white, black/African American, etc.)	Hispanic group imputed
802	Combination race and non-Hispanic country	Other Hispanic/Latino
985	Bad Data/"Mixed"	Hispanic/Latino group imputed
994	Unknown/"Don't Know"	Hispanic/Latino group imputed
997	American or "All of Them"	Hispanic/Latino group imputed

* This code caused the Hispanic/Latino indicator to be edited to a "no." Codes that caused the Hispanic/Latino indicator to be edited to a "yes" are listed in Table C.1.

Appendix D: Model Summaries

Appendix D: Model Summaries

D.1 Introduction

The tables in this appendix list the covariates used in all the imputation models that were run in the 2012 National Survey on Drug Use and Health (NSDUH). For each variable or set of variables to which the predictive mean neighborhood (PMN) imputation method was applied, two models were run: one to adjust the weights for item nonresponse (response propensity models) and a second to calculate predicted means. Imputation was usually done separately among age groups. Therefore, most of the tables within this appendix display only one age group at a time.

The models for the demographic variables are presented in Section D.3 and the models for the drug variables are presented in Section D.4. With the exception of the lifetime usage models, separate tables are provided in Section D.4 for each drug age group combination. Tables that present the models for each age group for the household composition variables, which are derived from the questionnaire roster items, are provided in Section D.5. Section D.6 presents the models for the income variables and Section D.7 presents the models for the health insurance variables. Section D.7 also presents the models for both the "Old Method" and the "Constituent Variables Method," used to create the final imputation-revised health insurance variables. Chapter 9 provides a more detailed description of these two methods. Section D.8 presents the models for the roster pair variables.

The definition of terms and variables in the models of the various sections can be found in the chapters that correspond to those sections as follows: Section D.3 (Chapters 3 & 4), Section D.4 (Chapters 5 & 6), Section D.5 (Chapter 7), Section D.6 (Chapter 8), Section D.7 (Chapter 9), and Section D.8 (Chapter 10).

In the tables, the variable "age" is the mean-centered age, where the age was "centered" by subtracting the mean age and where the mean was calculated across all respondents within the age group who were used to build the given model. The variables "age squared" and "age cubed" represent the square and cube, respectively, of this mean-centered age variable. Also in the tables, when an asterisk "*" is given, it represents an interaction between two variables.

D.2 Screener and Segment-Level Variables

In the PMN procedure, statistical modeling was performed to adjust weights for item nonresponse and also to calculate predicted means in the imputation models. Descriptions of questionnaire-derived variables are described in the main body of this report, with the exception of screener and segment-level variables, which are described below. These variables were often used as covariates in both types of models for the PMN procedures.

D.2.1 Household Type

Household type was a three-level race/ethnicity variable based on screener data. It was created by recoding the race/ethnicity of the screening head of the household to one of three

levels: Hispanic/Latino, non-Hispanic/Latino black/African American, or non-Hispanic/Latino non-black/African American.

D.2.2 Census Region

Census region was a four-level geographic variable recoded from the respondent's State of residence. The four levels were Northeast, Midwest, South, and West. The Northeast includes Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont. The Midwest consists of Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia compose the South. Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming compose the West.

D.2.3 Population Density

The population density variable classifies respondents according to their living situation, whether it be in a rural or urban area and, if urban, the size of the urban area. It was used to categorize segments where the respondents lived according to modified 2000 census data, which were adjusted to reflect population increases between census years by intercensal projections by Claritas, Inc.[1] This variable had five levels: segment in core-based statistical area (CBSA)[2] with 1 million or more persons; segment in CBSA with 250,000 to 999,999 persons; segment in CBSA with fewer than 250,000 persons; segment not in CBSA and not in rural area; and segment not in CBSA and in rural area.

D.2.4 Percentage Hispanic/Latino in Segment

The "Percentage Hispanic/Latino in Segment" variable was used to categorize segments according to the concentration of Hispanics/Latinos in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 20 percent, 20 to less than 71 percent, and 71 percent or more.

D.2.5 Percentage Owner Occupied in Segment

The "Percentage Owner Occupied in Segment" variable was used to categorize segments according to the concentration of owner-occupied households in the segments in which the respondents lived, using the adjusted 2000 census data. It was used as a surrogate for income because wealthy segments tend to have many homeowners, while poor segments tend to have many renters. It had three levels: less than 10 percent, 10 to less than 50 percent, and 50 percent or more.

---

[1] Claritas, Inc. is a market research firm headquartered in San Diego, California.

[2] CBSAs, developed in response to standards put forth by the Office of Management and Budget (OMB), are metropolitan and micropolitan areas that were designated using data from the 2000 census. More information about CBSAs can be retrieved from http://www.census.gov/hhes/www/housing/resseg/cbsa.html.

D.2.6 Percentage Black/African American in Segment

The "Percentage Black/African American in Segment" variable was used to categorize segments according to the concentration of black/African-American households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 10 percent, 10 to less than 40 percent, and 40 percent or more.

D.2.7 Percentage Asian/Other Pacific Islander in Segment

The "Percentage Asian/Other Pacific Islander in Segment" variable was used to categorize segments according to the concentration of Asian/Other Pacific Islander households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 5 percent, 5 to less than 10 percent, and 10 percent or more.

D.2.8 Percentage American Indian/Alaska Native in Segment

The "Percentage American Indian/Alaska Native in Segment" variable was used to categorize segments according to the concentration of American Indian/Alaska Native households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 1 percent, 1 to less than 3 percent, and 3 percent or more.

D.2.9 Percentage Non-Hispanic/Latino Black/African American in Segment

The "Percentage Non-Hispanic/Latino Black/African American in Segment" variable was used to categorize segments according to the concentration of non-Hispanic/Latino black/African-American households in the segments in which the respondents lived, using the adjusted 2000 census data. It had three levels: less than 10 percent, 10 percent to less than 50 percent, and 50 percent or more.

D.3 Demographic Variables

For justifications of the aggregation of age groups for certain imputation steps, see Chapters 3 and 4.

Exhibit D.1 Definitions of Levels for Variables

Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
Race/Ethnicity of Householder
B: Black/African American, H: Hispanic/Latino, W: White[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Percentage Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Percentage American Indian/Alaska Native in Segment
H: ≥ 3%, M: [1,3)%, L: < 1%[1]
Percentage Asian/Other Pacific Islander in Segment
H: ≥ 10%, M: [5,10)%, L: < 5%[1]

Exhibit D.1 Definitions of Levels for Variables (continued)

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, O: Not Hispanic/Latino and Other, H: Hispanic/Latino

Race/Hispanic Recode (5 Levels)
W: White Only,[1] B: Black/African American Only, AI: American Indian/Alaska Native Only, A: Asian/Other Pacific Islander Only, MR: Multiple Race

Race/Hispanic Recode (8 Levels)
W: Not Hispanic/Latino and White,[1] B: Not Hispanic/Latino and Black/African American, A: Not Hispanic/Latino and Asian/Other Pacific Islander, AI: Not Hispanic/Latino and American Indian/Alaska Native, MR: Not Hispanic/Latino and Multiple Race, PR: Hispanic/Latino and Puerto Rican, M: Hispanic/Latino and Mexican, O: Not Hispanic/Latino and Other

Population Density
LC: Segment in a CBSA with 1 Million or More Persons, MC: Segment in a CBSA from 250,000 to 999,999 Persons, SC: Segment in a CBSA with Fewer than 250,000 Persons, NC: Segment Not in a CBSA and Not in a Rural Area, R: Segment Not in a CBSA and in a Rural Area[1]

MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Gender
M: Male, F: Female[1]

Age Category
Y: 12-17, T: 18-25, A1: 26-34, A2: 35-49, A3: 50+[1]

Age Category * Gender
MY: Male 12-17, MT: Male 18-25, MA1: Male 26-34, MA2: Male 35-49, MA3: Male 50+,[1] FY: Female 12-17,[1] FT: Female 18-25,[1] FA1: Female 26-34,[1] FA2: Female 35-49,[1] FA3: Female 50+[1]

Race/Hispanic Recode (8 Levels) * Gender
MW: Male Not Hispanic/Latino and White,[1] MB: Male Not Hispanic/Latino and Black/African American, MA: Male Not Hispanic/Latino and Asian/Other Pacific Islander, MAI: Male Not Hispanic/Latino and American Indian/Alaska Native, MMR: Male Not Hispanic/Latino and Multiple Race, MPR: Male Hispanic/Latino and Puerto Rican, MM: Male Hispanic/Latino and Mexican, MO: Male Not Hispanic/Latino and Other, FW: Female Not Hispanic/Latino and White,[1] FB: Female Not Hispanic/Latino and Black/African American,[1] FA: Female Not Hispanic/Latino and Asian/Other Pacific Islander,[1] FAI: Female Not Hispanic/Latino and American Indian/Alaska Native,[1] FMR: Female Not Hispanic/Latino and Multiple Race,[1] FPR: Female Hispanic/Latino and Puerto Rican,[1] FM: Female Hispanic/Latino and Mexican,[1] FO: Female Not Hispanic/Latino and Other[1]

Imputation-Revised Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Imputation-Revised Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

Imputation-Revised Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]

Imputation-Revised Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced or Separated, NM: Never Been Married[1]

Imputation-Revised Hispanic/Latino Origin Indicator
H: Hispanic/Latino, NH: Not Hispanic/Latino[1]

NOTE: An asterisk "*" represents an interaction between two variables.

CBSA = core-based statistical area; MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.1 Summaries for Response Propensity Models

Imputation Step	Variables Included in Response Propensity Model
Marital Status	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Population Density (LC, MC, SC, NC); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2)
Race 12-17	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M)
Race 18-25	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Imputation-Revised Marital Status (3 Levels) (M, WM)
Race 26+	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Age Category (A1, A2); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Hispanic/Latino Origin 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 18-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 26+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Age Category (A1, A2)
Hispanic/Latino Group	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2)
Education Level 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M)
Education Level 18+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (T, A1, A2); Age Category * Gender (MT, MA1, MA2)

Table D.1 Summaries for Response Propensity Models (continued)

Imputation Step	Variables Included in Response Propensity Model
Employment Status 12-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (Y); Age Category * Gender (MY); **Imputation-Revised Education Level (4 Levels) (**L, **HS**, **SC)**
Employment Status 26+	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age Category (A1, A2); Age Category * Gender (MA1, MA2); Imputation-Revised Education Level (4 Levels) (L, HS, SC)
Born in US 12-17	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M)
Born in US 18-25	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Born in US 26+	**Census Region (**N, **M**, S); **Percentage Owner Occupied in Segment (**H, **M)**; **Race/Hispanic Recode (8 Levels) (**B, **A**, **AI**, **MR**, **PR**, M, O**)**; **MSA (R**, **SM)**; Gender (M); **Age Category (A1**, **A2)**; **Age Category * Gender (MA1**, **MA2)**; **Imputation-Revised Education Level (4 Levels) (**L, HS, **SC)**; **Imputation-Revised Employment Status (**FT, **PT**, **UN)**; **Imputation-Revised Marital Status (3 Levels) (**M, **WM)**
Age of Entry	Census Region (N, M, S); Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age Category (Y, T, A1, A2); Age Category * Gender (MY, MT, MA1, MA2); Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO); Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.1 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.2 Summaries for Predictive Mean Models

Imputation Step	Variables Included in Predictive Mean Model
Marital Status	Census Region (N, M, S); Age; **Age Squared**; **Age Cubed**; **Percentage Owner Occupied in Segment (**H, **M)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Black/African American in Segment (H, M); **Percentage American Indian/Alaska Native in Segment (**H, **M)**; **Percentage Asian/Other Pacific Islander in Segment (**H, **M)**; **Population Density (LC**, **MC**, **SC**, **NC)**; Gender (M); Age * Gender
Race 12-17	Census Region (N, M, S); **Race/Ethnicity of Householder (B**, **H)**; Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M)
Race 18-25	Census Region (N, M, S); **Race/Ethnicity of Householder (B**, **H)**; Age; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Imputation-Revised Marital Status (3 Levels) (M, WM)
Race 26+	Census Region (N, M, S); **Race/Ethnicity of Householder (B**, **H)**; Age; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Hispanic/Latino Origin 12-17	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR)
Hispanic/Latino Origin 18-25	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Imputation-Revised Marital Status (3 Levels) (M, WM)
Hispanic/Latino Origin 26+	Census Region (N, M, S); Race/Ethnicity of Householder (B, H); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Hispanic/Latino Group	Census Region (N, M, S); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (5 Levels) (B, AI, A, MR); Gender (M); Age * Gender; Age Squared * Gender
Education Level 12-17	Census Region (N, M, S); **Age**; **Age Squared**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M)

Table D.2 Summaries for Predictive Mean Models (continued)

Imputation Step	Variables Included in Predictive Mean Model
Education Level 18+	Census Region (N, M, S); Age; Age Squared; Age Cubed; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Employment Status 12-25	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (4 Levels) (L, HS, SC)
Employment Status 26+	Census Region (N, M, S); Age; **Age Squared**; **Age Cubed**; Percentage Owner Occupied in Segment (H, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage American Indian/Alaska Native in Segment (H, M); Percentage Asian/Other Pacific Islander in Segment (H, M); Race/Hispanic Recode (4 Levels) (B, O, H); Gender (M); Age * Gender; **Age Squared * Gender**; Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Marital Status (4 Levels) (M, W, DS)
Born in US 12-17	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender
Born in US 18-25	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Born in US 26+	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)
Age of Entry	Census Region (N, M, S); Age; Age Squared; Percentage Owner Occupied in Segment (H, M); Race/Hispanic Recode (8 Levels) (B, A, AI, MR, PR, M, O); MSA (R, SM); Gender (M); Age * Gender; Age Squared * Gender; Race/Hispanic Recode (8 Levels) * Gender (MB, MA, MAI, MMR, MPR, MM, MO); Imputation-Revised Education Level (4 Levels) (L, HS, SC); Imputation-Revised Employment Status (FT, PT, UN); Imputation-Revised Marital Status (3 Levels) (M, WM)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.1 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

D.4 Drug Variables

Exhibit D.2 Definitions of Levels for Variables

Age Category
A1: 26-34, A2: 35-49, A3: 50+[1]
Gender
M: Male, F: Female[1]
Race/Hispanic Recode (2 Levels)[2]
W: Not Hispanic/Latino and White Only, O: Hispanic/Latino or Not White Only[1]
Race/Hispanic Recode (4 Levels)[2]
W: Not Hispanic/Latino and White Only,[1] B: Not Hispanic/Latino and Black/African American Only, H: Hispanic/Latino, O: Not Hispanic/Latino and Other[1]
Gender * Race/Hispanic Recode (4 Levels)[2]
MW: Male Not Hispanic/Latino and White Only,[1] MB: Male Not Hispanic/Latino and Black/African American Only, MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other,[1] FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other[1]
Age * Race/Hispanic Recode (4 Levels)[2]
W: Age * Not Hispanic/Latino and White Only,[1] B: Age * Not Hispanic/Latino and Black/African American Only, H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other[1]
Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other
Marital Status (2 Levels)
M: Married, NM: Not Married[1]
Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]
Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
State Rank
L: Low State Rank (lowest tertile), M: Middle State Rank (middle tertile), H: High State Rank (highest tertile)[1]
Cigarette Lifetime Indicator
Y: Yes, N: No[1]
Intermediate Drug-Specific Past Month Indicator
Y: Yes, N: No[1]
Intermediate Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Imputation-Revised Drug-Specific Lifetime Indicator
Y: Yes, N: No[1]
Imputation-Revised Pipe Recency
PM: Past Month, LF: Lifetime but Not Past Month, NU: Lifetime Nonuser[1]
Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels)
PM: Past Month, PY: Past Year but Not Past Month, LF: Lifetime but Not Past Year[1]
Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels)
PM: Past Month, PY: Past Year but Not Past Month, LF: Lifetime but Not Past Year, NU: Lifetime Nonuser[1]

Exhibit D.2 Definitions of Levels for Variables (continued)

Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) PM: Past Month, PY: Past Year but Not Past Month, P3Y: Past 3 Years but Not Past Year, LF: Lifetime but Not Past 3 Years[1] **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels)** PM: Past Month, PY: Past Year but Not Past Month, P3Y: Past 3 Years but Not Past Year, LF: Lifetime but Not Past 3 Years, NU: Lifetime Nonuser[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

[2] For the vast majority of drug models, the reference cell for the four-level race/Hispanic recode was "Not Hispanic/Latino and White Only." For a few models, the reference cell was "Not Hispanic/Latino and Other." The atypical models were those for which statisticians were encouraged to include the "Not Hispanic/Latino and White Only" level as a good covariate, because of past performance. In general, the choice of "Not Hispanic/Latino and White Only" as the reference cell is slightly preferable because it facilitates comparisons to the most populous category in the United States. Interactions involving the race/Hispanic recode are handled similarly.

Table D.3 Lifetime Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12-17	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)
18-25	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)
26+	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Cigarette Lifetime Indicator (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.4 Cigarettes: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y)
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y) Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)

Table D.4 Cigarettes: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)
Ever Daily Used	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.5 Cigarettes: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)

Table D.5 Cigarettes: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

Table D.5 Cigarettes: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.6 Cigarettes: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	N/A	N/A
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)
12-Month Frequency	N/A	N/A
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)	Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)

Table D.6 Cigarettes: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y)
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

Table D.6 Cigarettes: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Ever Daily Used	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Age at First Use for Cigarettes; Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 30-Day Frequency for Cigarettes; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B, H, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO)**; **Age * Race/Hispanic Recode (4 Levels) (B, H, O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B, H, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO)**; **Age * Race/Hispanic Recode (4 Levels) (B, H, O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), **Heroin (Y)**, and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco Daily Users: **Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)** Chewing Tobacco Nondaily Users: Age; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); Imputation-Revised Drug-Specific Lifetime Indicator for Alcohol (Y), Heroin (Y), and Hallucinogens (Y)

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff Daily Users: Age; Age Squared; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)** Snuff Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.7 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Chewing Tobacco: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Chewing Tobacco: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MH, MO**)**; **Age * Race/Hispanic Recode (4 Levels) (B**, H, O**)**; Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Chewing Tobacco: Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), Crack (Y), **Heroin (Y)**, and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**	Chewing Tobacco Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Chewing Tobacco Nondaily Users: Age; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (**N, M, **S)**; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y)
	Snuff: Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff Daily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (**MW, **MB**, **MH)**; **Age * Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF) Snuff Nondaily Users: Age; Age Squared; Age Cubed; **Gender (M)**; Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.8 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender
Recency	Smokeless Tobacco: Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Smokeless Tobacco: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
	Chewing Tobacco: **Age Category (**A1, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (**R, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Chewing Tobacco: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Snuff: Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Snuff: **Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (**FT, PT, **UN)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), Inhalants (Y), **Marijuana (Y)**, Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	N/A	N/A

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Chewing Tobacco: **Age Category (**A1, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Pipes (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Chewing Tobacco Daily Users: **Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF) Chewing Tobacco Nondaily Users: Age; Race/Hispanic Recode (4 Levels) (B, H, O); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
	Snuff: Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); **MSA (R**, **SM)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Pipes (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Snuff Daily Users: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.9 Smokeless Tobacco (Chewing Tobacco and Snuff): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
		Snuff Nondaily Users: Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (**M, W, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; Employment Status (FT, PT, UN); **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (**L, **M)**; Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), **Marijuana (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Pipes (Y)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Smokeless Tobacco (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, and Daily Cigarette Use; Imputation-Revised Drug-Specific 30-Day Frequency for Chewing Tobacco, and Snuff; Imputation-Revised Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Snuff (PM, PY, P3Y, LF), and Chewing Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.10 Cigars: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.10 Cigars: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.11 Cigars: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A

Table D.11 Cigars: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.12 Cigars: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Chewing Tobacco (Y), and Snuff (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pipes (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A

Table D.12 Cigars: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator for** Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), **Heroin (Y)**, and Alcohol (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cigars; Imputation-Revised Drug-Specific Lifetime Indicator for Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Alcohol (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigars (PM, PY, P3Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.13 Pipes: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	No Model Used; No Nonrespondents	No Model Used; No Nonrespondents
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area; N/A = not applicable.

Table D.14 Pipes: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.15 Pipes: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Snuff (Y), and Chewing Tobacco (Y)
Recency	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Alcohol (Y)**, Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), **Heroin (Y)**, and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.16 Alcohol: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.16 Alcohol: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Binge Drinking Frequency	N/A	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.16 Alcohol: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.17 Alcohol: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Binge Drinking Frequency	N/A	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.17 Alcohol: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.18 Alcohol: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.18 Alcohol: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Alcohol (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.18 Alcohol: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Binge Drinking Frequency	N/A	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Intermediate Alcohol 30-Day Frequency; Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.18 Alcohol: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Alcohol; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Alcohol; Imputation-Revised Drug-Specific Lifetime Indicator for Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Marijuana (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.19 Inhalants: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.19 Inhalants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Census Region (**N, M, **S)**; **MSA (**R, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; **Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and **Hallucinogens (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; Age * Gender; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; **Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Hallucinogens (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**

Table D.19 Inhalants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.20 Inhalants: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.20 Inhalants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MH**, **MO)**; **Marital Status (3 Levels) (M**, **WM)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; Employment Status (FT, PT, UN); **Census Region (N**, **M**, **S)**; MSA (R, SM); State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Census Region (N**, **M**, **S)**; MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for** Marijuana (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), Heroin (Y), and **Hallucinogens (Y)**; **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.20 Inhalants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (**R, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; **Imputation-Revised Drug-Specific Lifetime Indicator for** Marijuana (Y), Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and Hallucinogens (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Race/Hispanic Recode (4 Levels) (B, H, O); Education Level (4 Levels) (L, HS, SC); Census Region (N, M, S); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants

Table D.20 Inhalants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.21 Inhalants: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Marijuana (Y), Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.21 Inhalants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Age Category (**A1, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Marijuana (Y), **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific Past Month Indicator for Inhalants (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Marijuana (Y), **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Hallucinogens (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Race/Hispanic Recode (2 Levels) (W); Education Level (4 Levels) (L); Employment Status (FT, PT, UN); Intermediate Drug-Specific Past Month Indicator for Inhalants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), and Pain Relievers (Y)

Table D.21 Inhalants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants; **Imputation-Revised Drug-Specific Lifetime Indicator for** Marijuana (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Hallucinogens (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Employment Status (FT**, **PT**, **UN)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Inhalants**; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), and Crack (Y)

Table D.21 Inhalants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Inhalants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Inhalants; Imputation-Revised Drug-Specific Lifetime Indicator for Marijuana (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Hallucinogens (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Inhalants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.22 Marijuana: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.22 Marijuana: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.22 Marijuana: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.23 Marijuana: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.23 Marijuana: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.23 Marijuana: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.24 Marijuana: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.24 Marijuana: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Marijuana (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.24 Marijuana: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Marijuana; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Marijuana; Imputation-Revised Drug-Specific Lifetime Indicator for Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Pain Relievers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.25 Hallucinogens: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (**PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.25 Hallucinogens: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (**PM, PY, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Pain Relievers (Y), Stimulants (Y), **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, and **Inhalants (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (**R, **SM)**; State Rank (L, M); **Age * Gender**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Pain Relievers (Y), **Stimulants (Y)**, Sedatives (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Tranquilizers (Y); **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, and **Inhalants (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**

Table D.25 Hallucinogens: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.26 Hallucinogens: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.26 Hallucinogens: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.26 Hallucinogens: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.26 Hallucinogens: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.27 Hallucinogens: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.27 Hallucinogens: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (**N, M, **S)**; MSA (R, SM); **State Rank (**L, **M)**; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, H, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (**N, M, **S)**; MSA (R, SM); **Intermediate Drug-Specific Past Month Indicator for Hallucinogens (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Pain Relievers (Y), **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, Crack (Y), Heroin (Y), and **Tranquilizers (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**, Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (4 Levels) (B, H, O); Education Level (4 Levels) (L, HS, SC); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), and Stimulants (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF)

Table D.27 Hallucinogens: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Pain Relievers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and Tranquilizers (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Race/Hispanic Recode (2 Levels) (W); Marital Status (2 Levels) (M); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Hallucinogens; Imputation-Revised Pipe Recency (PM)

Table D.27 Hallucinogens: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Hallucinogens; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Hallucinogens; Imputation-Revised Drug-Specific Lifetime Indicator for Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Tranquilizers (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Hallucinogens (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.28 Pain Relievers: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.28 Pain Relievers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.28 Pain Relievers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.29 Pain Relievers: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.29 Pain Relievers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.29 Pain Relievers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.30 Pain Relievers: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.30 Pain Relievers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Pain Relievers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.30 Pain Relievers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Pain Relievers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Stimulants (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Pain Relievers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.31 Tranquilizers: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.31 Tranquilizers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**
12-Month Frequency	**Gender (M)**; Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for** Stimulants (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Sedatives (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), **Pain Relievers (**PM, PY, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.31 Tranquilizers: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.32 Tranquilizers: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.32 Tranquilizers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.32 Tranquilizers: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.33 Tranquilizers: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.33 Tranquilizers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Cigars (Y)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y)**, **Cocaine (Y)**, Crack (Y), Heroin (Y), and Sedatives (Y); **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Tranquilizers (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.33 Tranquilizers: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Tranquilizers; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Sedatives (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.34 Stimulants: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (**MB, MH, **MO)**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
Recency: Past Month vs. Past Year Not Past Month	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MH**, **MO)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.34 Stimulants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and **Inhalants (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.34 Stimulants: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, PY, P3Y, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.35 Stimulants: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.35 Stimulants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.35 Stimulants: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	N/A	N/A
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.36 Stimulants: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.36 Stimulants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; Employment Status (FT, PT, UN); **Census Region (N**, **M**, **S)**; MSA (R, SM); State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (4 Levels) (**M, **W**, DS**)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for** Sedatives (Y), **Crack (Y)**, Heroin (Y), and Cocaine (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, PY, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, Tranquilizers (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Intermediate Drug-Specific Past Month Indicator for Stimulants (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Heroin (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Marijuana (PM); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (4 Levels) for Cigarettes (PM)
30-Day Frequency	N/A	N/A

Table D.36 Stimulants: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Stimulants; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Sedatives (Y), Crack (Y), Heroin (Y), Daily Cigarette Use (Y), and Cocaine (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Stimulants (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.37 Sedatives: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MH**, **MO)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.37 Sedatives: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (**MB, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (**B, H, **O)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, SM); Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Census Region (**N, M, **S)**; MSA (R, SM); **State Rank (**L, **M)**; **Age * Gender**; **Intermediate Drug-Specific Past Month Indicator for Sedatives (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cocaine (Y), **Heroin (Y)**, and **Crack (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**
30-Day Frequency	N/A	N/A

Table D.37 Sedatives: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); **MSA (**R, **SM)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, Heroin (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (PM**, **PY**, **LF)**, Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and **Inhalants (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.38 Sedatives: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (3 Levels) (M**, **WM)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (3 Levels) (**M, **WM)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.38 Sedatives: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (3 Levels) (M**, **WM)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; MSA (R, SM); **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Stimulants (Y), **Cocaine (Y)**, Crack (Y), Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (3 Levels) (M**, **WM)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Cocaine (Y)**, **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**
12-Month Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (**R, **SM)**; **Intermediate Drug-Specific Past Month Indicator for Sedatives (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Cocaine (Y), Heroin (Y), and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, LF), and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (3 Levels) (M**, **WM)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Intermediate Drug-Specific Past Month Indicator for Sedatives (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y)**, **Heroin (Y)**, and **Crack (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**

Table D.38 Sedatives: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	N/A	N/A
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Census Region (**N, M, **S)**; MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, PY, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), and **Inhalants (PM**, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.39 Sedatives: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; MSA (R, SM); **State Rank (L**, **M)**; Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, **Cocaine (Y)**, Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.39 Sedatives: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Heroin (Y)**, and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Cocaine (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	No Model Used; No Nonrespondents	Gender (M); Marital Status (2 Levels) (M); Intermediate Drug-Specific Past Month Indicator for Sedatives (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
30-Day Frequency	N/A	N/A

Table D.39 Sedatives: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Sedatives; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific Lifetime Indicator for Cocaine (Y), Heroin (Y), Daily Cigarette Use (Y), and Crack (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Sedatives (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.40 Cocaine (and Crack): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)** Crack: **Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (B, H, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO)**; **Age * Race/Hispanic Recode (4 Levels) (B, H, O)**; **Census Region (N, M, S)**; **MSA (R, SM)**; **State Rank (L, M)**; Cigarette Lifetime Indicator (Y); **Age * Gender**; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), **Sedatives (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**
Recency	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L, M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), Sedatives (Y), **Crack (Y)**, Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM, PY, LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y, LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MH, MO**)**; **Age * Race/Hispanic Recode (4 Levels) (B**, H, O**)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF)**, and **Marijuana (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y, LF)**

Table D.40 Cocaine (and Crack): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
12-Month Frequency	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MH, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, H, **O)**; Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); Age * Gender; **Intermediate Drug-Specific Past Month Indicator for Cocaine (Y)**; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (**PM, PY, **LF)**, Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and **Smokeless Tobacco (**PM, PY, P3Y, **LF)**
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; Census Region (N, M, S); **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigars (**PM, PY, **P3Y**, **LF)**

Table D.40 Cocaine (and Crack): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); **Imputation-Revised Drug-Specific Lifetime Indicator for** Crack (Y), and **Heroin (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, Marijuana (PM, PY, LF), **Hallucinogens (**PM, PY, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (**PM, PY, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Age * Gender**; **Age Squared * Gender**; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; **Drug-Specific Imputation-Revised Age at First Use for** Cigarettes, Cigars, **Alcohol**, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (**PM, **PY)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.41 Cocaine (and Crack): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y) Crack: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)

Table D.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Intermediate Drug-Specific Past Month Indicator for Cocaine (Y); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

Table D.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, H, O); **Census Region (N**, **M**, **S)**; **MSA (**R, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, Tranquilizers (PM, PY, LF), **Stimulants (**PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (3 Levels) (M**, **WM)**; **Education Level (4 Levels) (**L, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Crack (Y), and **Heroin (Y)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (**PM, PY, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (**PM, PY, **LF)**, **Stimulants (**PM, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**

Table D.41 Cocaine (and Crack): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.42 Cocaine (and Crack): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Cocaine: Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y) Crack: Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)

Table D.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; **Census Region (N**, M, S**)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), **Pain Relievers (Y)**, Tranquilizers (Y), Stimulants (Y), Sedatives (Y), **Crack (Y)**, **Heroin (Y)**, and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific Past Month Indicator for Cocaine (Y)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and Heroin (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (**PM, PY, **LF)**, **Tranquilizers (**PM, **PY**, **LF)**, **Stimulants (**PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; **Intermediate Drug-Specific Past Month Indicator for Cocaine (Y)**; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), **Sedatives (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for** Cigarettes (PM, PY, P3Y, LF), **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, PY, **P3Y**, **LF)**

Table D.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, A2**)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, O**)**; **Census Region (N**, M, **S)**; MSA (R, SM); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine; **Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)**, and **Heroin (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, Marijuana (PM, PY, LF), **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (**PM, PY, **LF)**, **Tranquilizers (**PM, **PY**, **LF)**, **Stimulants (**PM, PY, **LF)**, **Sedatives (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, P3Y, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	**Age**; Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MH, **MO)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Cocaine**; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y)

Table D.42 Cocaine (and Crack): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Cocaine; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Cocaine; Imputation-Revised Drug-Specific Lifetime Indicator for Crack (Y), Daily Cigarette Use (Y), and Heroin (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.43 Heroin: Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); MSA (R, SM); **State Rank (L**, **M)**; **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for Cigars (Y)**, **Alcohol (Y)**, **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, **Chewing Tobacco (Y)**, **Snuff (Y)**, and **Pipes (Y)**
Recency: Past Year vs. Not Past Year	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Census Region (N**, **M**, S)**; MSA (R, SM); **State Rank (**L, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, **Crack (Y)**, and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, S)**; MSA (R, SM); **State Rank (**L, **M)**; Age * Gender; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, Sedatives (Y), **Cocaine (Y)**, Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, P3Y, **LF)**

Table D.43 Heroin: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, PY, **P3Y**, **LF)**	No Model Used; No Nonrespondents
12-Month Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; MSA (R, SM); **Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Age; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Intermediate Drug-Specific Past Month Indicator for Heroin (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (**PM, PY, **LF)**
30-Day Frequency	No Model Used; No Nonrespondents	No Model Used; No Nonrespondents

Table D.43 Heroin: Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, M, **S)**; **MSA (**R, **SM)**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Age Squared * Gender**; **Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin**; **Drug-Specific Imputation-Revised Age at First Use for Cigarettes**, **Cigars**, **Alcohol**, **Inhalants**, **Marijuana**, **Hallucinogens**, **Pain Relievers**, **Tranquilizers**, **Stimulants**, **Sedatives**, **Cocaine**, **Crack**, **Daily Cigarette Use**, and **Smokeless Tobacco**; **Imputation-Revised Drug-Specific 30-Day Frequency for Heroin**; **Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y)**; **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, **Marijuana (**PM, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (**PM, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**

NOTE: Boldface of variables and levels indicate\12nsduh\table\tabshl\. that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.44 Heroin: Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Cigarette Lifetime Indicator (Y)**; Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), **Crack (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.44 Heroin: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (3 Levels) (**M, **WM)**; **Education Level (4 Levels) (**L, HS, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (**R, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (3 Levels) (**M, **WM)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Stimulants (Y), Sedatives (Y), Cocaine (Y), Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (**B, **H**, O); **Census Region (N**, M, **S)**; **MSA (**R, **SM)**; Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (**PM, **PY**, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (**PM, PY, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (**PM, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Marital Status (3 Levels) (M, WM); Intermediate Drug-Specific Past Month Indicator for Heroin (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (PM, PY, LF), Crack (PM, PY, LF), and Tranquilizers (PM, PY, LF)

Table D.44 Heroin: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Gender (M)**; **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Census Region (N**, M, S**)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), **Hallucinogens (**PM, PY, **LF)**, **Pain Relievers (PM**, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and Inhalants (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (2 Levels) (W); **MSA (R)**; Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Tranquilizers (PM), and Crack (PM)

Table D.44 Heroin: Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; Census Region (N, M, S); **MSA (R**, **SM)**; **Imputation-Revised Pipe Recency (**PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, **Marijuana (PM**, **PY**, **LF)**, Hallucinogens (PM, PY, LF), **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (**PM, PY, **LF)**, **Sedatives (PM**, **PY**, **LF)**, Cocaine (PM, PY, LF), **Crack (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (**PM, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (**PM, PY, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Age Squared * Gender; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; Drug-Specific Imputation-Revised Age at First Use for Cigarettes, Cigars, Alcohol, Inhalants, Marijuana, Hallucinogens, Pain Relievers, Tranquilizers, Stimulants, Sedatives, Cocaine, Crack, Daily Cigarette Use, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Heroin; Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y); Imputation-Revised Pipe Recency (PM, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), Pain Relievers (PM, PY, LF), Tranquilizers (PM, PY, LF), Stimulants (PM, PY, LF), Sedatives (PM, PY, LF), Cocaine (PM, PY, LF), Crack (PM, PY, LF), and Inhalants (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF), Cigars (PM, PY, P3Y, LF), and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

Table D.45 Heroin: Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, **Marijuana (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, **Crack (Y)**, Chewing Tobacco (Y), Snuff (Y), and **Pipes (Y)**
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, MH, **MO)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; **Education Level (4 Levels) (**L, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), **Pipes (Y)**, **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, Cocaine (Y), Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), Sedatives (Y), **Cocaine (Y)**, Crack (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**

Table D.45 Heroin: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	**Age Category (A1**, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; **Education Level (4 Levels) (**L, HS, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, **Inhalants (Y)**, **Hallucinogens (Y)**, **Pain Relievers (Y)**, Tranquilizers (Y), **Stimulants (Y)**, **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (**FT, PT, **UN)**; **Census Region (**N, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Age * Gender**; **Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y)**, **Pipes (Y)**, Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), **Tranquilizers (Y)**, Stimulants (Y), **Sedatives (Y)**, **Cocaine (Y)**, Crack (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and **Marijuana (**PM, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**
12-Month Frequency	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (**N, **M**, S); **MSA (R**, **SM)**; Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (**PM, **PY**, **P3Y**, **LF)**	Race/Hispanic Recode (2 Levels) (W); Marital Status (2 Levels) (M); Intermediate Drug-Specific Past Month Indicator for Heroin (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Marijuana (**PM, **PY**, **LF)**, and **Cocaine (**PM, **PY**, **LF)**

Table D.45 Heroin: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
30-Day Frequency	**Age Category (A1**, **A2)**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin**; **Imputation-Revised Pipe Recency (PM**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (**PM, PY, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, **Pain Relievers (**PM, **PY**, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	Gender (M); Employment Status (FT); Intermediate Drug-Specific 12-Month Frequency, Domain Restricted to Past Month Users for Heroin; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Cocaine (PM)

Table D.45 Heroin: Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Age at First Use	**Age Category (**A1, **A2)**; Gender (M); **Race/Hispanic Recode (4 Levels) (**B, **H**, **O)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; **Imputation-Revised Pipe Recency** (PM, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM**, **PY**, **LF)**, **Marijuana (PM**, **PY**, **LF)**, **Hallucinogens (PM**, **PY**, **LF)**, Pain Relievers (PM, PY, LF), **Tranquilizers (PM**, **PY**, **LF)**, **Stimulants (PM**, **PY**, **LF)**, **Sedatives (PM**, **PY**, **LF)**, **Cocaine (PM**, **PY**, **LF)**, **Crack (PM**, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM**, **PY)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and **Smokeless Tobacco (PM**, **PY**, **P3Y**, **LF)**	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, MO); **Age * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Age Squared * Race/Hispanic Recode (4 Levels) (B**, **H**, **O)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, HS, **SC)**; Employment Status (FT, PT, UN); **Census Region (N**, M, **S)**; **MSA (R**, **SM)**; State Rank (L, M); **Age * Gender**; **Age Squared * Gender**; Imputation-Revised Drug-Specific 12-Month Frequency, Domain Restricted to Lifetime Users for Heroin; **Drug-Specific Imputation-Revised Age at First Use for Cigarettes**, **Cigars**, Alcohol, **Inhalants**, Marijuana, Hallucinogens, **Pain Relievers**, **Tranquilizers**, Stimulants, **Sedatives**, **Cocaine**, **Crack**, **Daily Cigarette Use**, and Smokeless Tobacco; Imputation-Revised Drug-Specific 30-Day Frequency for Heroin; **Imputation-Revised Drug-Specific Lifetime Indicator for Daily Cigarette Use (Y)**; Imputation-Revised Pipe Recency (PM, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), Marijuana (PM, PY, LF), Hallucinogens (PM, PY, LF), **Pain Relievers (**PM, PY, **LF)**, **Tranquilizers (PM**, **PY**, **LF)**, Stimulants (PM, PY, LF), **Sedatives (PM**, **PY**, **LF)**, Cocaine (PM, PY, LF), **Crack (**PM, **PY**, **LF)**, and **Inhalants (PM**, **PY**, **LF)**; Imputation-Revised Drug-Specific Recency, Domain Restricted to Lifetime Users (3 Levels) for Heroin (PM, PY); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**, **Cigars (PM**, **PY**, **P3Y**, **LF)**, and Smokeless Tobacco (PM, PY, P3Y, LF)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.46 Stimulants (Core plus Noncore): Persons 12 to 17 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; **Intermediate Drug-Specific Lifetime Indicator for** Cigars (Y), Alcohol (Y), **Inhalants (Y)**, Marijuana (Y), **Hallucinogens (Y)**, **Pain Relievers (Y)**, **Tranquilizers (Y)**, Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.46 Stimulants (Core plus Noncore): Persons 12 to 17 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; Census Region (N, M, S); MSA (R, SM); State Rank (L, M); **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), **Hallucinogens (Y)**, Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), **Crack (Y)**, Heroin (Y), and **Cigars (Y)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for** Alcohol (PM, PY, LF), and **Marijuana (PM**, **PY**, **LF)**; **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM**, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.47 Stimulants (Core plus Noncore): Persons 18 to 25 Years Old

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.47 Stimulants (Core plus Noncore): Persons 18 to 25 Years Old (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

Table D.48 Stimulants (Core plus Noncore): Persons 26 Years or Older

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Lifetime	See Table D.3.	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Cigarette Lifetime Indicator (Y); Age * Gender; Intermediate Drug-Specific Lifetime Indicator for Cigars (Y), Alcohol (Y), Inhalants (Y), Marijuana (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Chewing Tobacco (Y), Snuff (Y), and Pipes (Y)
Recency: Past Year vs. Not Past Year	Age Category (A1, A2); Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (PM, PY, LF), and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, **PY**, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)

Table D.48 Stimulants (Core plus Noncore): Persons 26 Years or Older (continued)

Imputation Step	Variables Included in Response Propensity Model	Variables Included in Drug Model
Recency: Past Month vs. Past Year Not Past Month	Age Category (A1, A2); Gender (M); **Race/Hispanic Recode (4 Levels) (**B, H, **O)**; **Gender * Race/Hispanic Recode (4 Levels) (MB**, **MH**, **MO)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **State Rank (L**, **M)**; **Imputation-Revised Drug-Specific Lifetime Indicator for** Smokeless Tobacco (Y), Pipes (Y), **Inhalants (Y)**, Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), **Cocaine (Y)**, Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (**PM, **PY**, **P3Y**, **LF)**	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Age * Race/Hispanic Recode (4 Levels) (B, H, O); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); State Rank (L, M); Age * Gender; Imputation-Revised Drug-Specific Lifetime Indicator for Smokeless Tobacco (Y), Pipes (Y), Inhalants (Y), Hallucinogens (Y), Pain Relievers (Y), Tranquilizers (Y), Sedatives (Y), Cocaine (Y), Crack (Y), Heroin (Y), and Cigars (Y); **Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (3 Levels) for Alcohol (**PM, PY, **LF)**, and Marijuana (PM, PY, LF); Imputation-Revised Drug-Specific Recency, Domain Not Restricted to Lifetime Users (4 Levels) for Cigarettes (PM, PY, P3Y, LF)
12-Month Frequency	N/A	N/A
30-Day Frequency	N/A	N/A
Age at First Use	N/A	N/A

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.2 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area; N/A = not applicable.

D.5 Household Composition Variables

Exhibit D.3 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]
Race/Hispanic Recode (4 Levels)
H: Hispanic/Latino, O: Not Hispanic/Latino and Other, B: Not Hispanic/Latino and Black/African American Only, W: Not Hispanic/Latino and White Only[1]
Gender * Race/Hispanic Recode (4 Levels)
MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other, MB: Male Not Hispanic/Latino and Black/African American Only, MW: Male Not Hispanic/Latino and White Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FW: Female Not Hispanic/Latino and White Only[1]
Age * Race/Hispanic Recode (4 Levels)
H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other, B: Age * Not Hispanic/Latino and Black/African American Only, W: Age * Not Hispanic/Latino and White Only[1]
Age Squared * Race/Hispanic Recode (4 Levels)
H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other, B: Age Squared * Not Hispanic/Latino and Black/African American Only, W: Age Squared * Not Hispanic/Latino and White Only[1]
Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]
Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]
Imputation-Revised Family in Household
Y: Other Family Members in Household,[1] N: No Other Family Members in Household

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.49 Household Composition: Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household

Table D.49 Household Composition: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; **Imputation-Revised Number of Persons Older than 64 Years Old in Household**
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household

Table D.49 Household Composition: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H, M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

Table D.49 Household Composition: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.50 Household Composition: Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	Age; **Age Squared**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (3 Levels) (**M, **WM)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; Employment Status (FT, PT, UN); **Census Region (N**, M, S**)**; **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**; **Total Persons in Household (Screener)**	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household

Table D.50 Household Composition: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (3 Levels) (M**, **WM)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**; **Imputation-Revised Household Size**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household

Table D.50 Household Composition: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (3 Levels) (**M, **WM)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

Table D.50 Household Composition: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H, O, B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B)**; Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.51 Household Composition: Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (**MH, **MO**, **MB)**; Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household

Table D.51 Household Composition: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, **MB)**; Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (H, O, B); **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household

Table D.51 Household Composition: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; **Imputation-Revised Number of Persons Older than 64 Years Old in Household**; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)**

Table D.51 Household Composition: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; **Imputation-Revised Family in Household (N)**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; **Imputation-Revised Number of Persons Older than 64 Years Old in Household**; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.52 Household Composition: Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Household Size (TOTPEOP)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**; Total Persons in Household (Screener)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Total Persons in Household (Screener)
Number of Persons Younger than 18 in Household (KID17)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (**H, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; **Census Region (**N, **M**, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Number of Eligible 12 to 17 in Household (Screener)**	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; **Age * Race/Hispanic Recode (4 Levels) (**H, O, **B)**; Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; Number of Eligible 12 to 17 in Household (Screener)
Number of Persons Older than 64 in Household (HH65)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (**H, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; **Census Region (**N, **M**, S); **MSA (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Number of Persons Younger than 18 Years Old in Household

Table D.52 Household Composition: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Other Family Present in Household (FAMSKIP)	No Model Used; No Nonrespondents	No Model Used; No Nonrespondents
Number of Respondent's Family Members in Household including Foster Relationships (FMLYSIZE)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**; Imputation-Revised Household Size; Imputation-Revised Family in Household (N); **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; **Imputation-Revised Number of Persons Older than 64 Years Old in Household**
Number of Respondent's Family Members in Household Younger than 18 including Foster Relationships (KIDFMLY)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (L**, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (H**, **M)**; Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (H, O, B); Age Squared * Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**

Table D.52 Household Composition: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household excluding Foster Relationships (FAMSIZE)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (FT**, **PT**, **UN)**; **Census Region (N**, **M**, S); **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**; Imputation-Revised Household Size; Imputation-Revised Family in Household (N); **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); **Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)**	**Age**; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (H, O, B); Gender * Race/Hispanic Recode (4 Levels) (MH, MO, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)

Table D.52 Household Composition: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model
Number of Respondent's Family Members in Household Younger than 18 excluding Foster Relationships (KIDFAMSZ)	**Age**; **Age Squared**; Gender (M); **Race/Hispanic Recode (4 Levels) (H**, **O**, B); **Gender * Race/Hispanic Recode (4 Levels) (MH**, **MO**, MB); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (4 Levels) (**M, W, **DS)**; **Education Level (4 Levels) (**L, **HS**, **SC)**; **Employment Status (**FT, **PT**, **UN)**; **Census Region (N**, **M**, **S)**; **MSA (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**; Imputation-Revised Household Size; Imputation-Revised Family in Household (N); Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household; Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships); Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships); Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)	Age; Age Squared; Gender (M); **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; **Gender * Race/Hispanic Recode (4 Levels) (**MH, MO, **MB)**; Age * Gender; Age Squared * Gender; **Age * Race/Hispanic Recode (4 Levels) (**H, O, **B)**; **Age Squared * Race/Hispanic Recode (4 Levels) (**H, O, **B)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Household Size**; **Imputation-Revised Family in Household (N)**; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; **Imputation-Revised Number of Persons Older than 64 Years Old in Household**; **Imputation-Revised Number of Respondent's Family Members in Household (Excludes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members in Household (Includes Foster Relationships)**; **Imputation-Revised Number of Respondent's Family Members Younger than 18 in Household (Includes Foster Relationships)**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.3 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.
MSA = metropolitan statistical area.

D.6 Income Variables

Exhibit D.4 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only, B: Not Hispanic/Latino and Black/African American Only, H: Hispanic/Latino, O: Not Hispanic/Latino and Other[1]

Gender * Race/Hispanic Recode (4 Levels)
MW: Male Not Hispanic/Latino and White Only, MB: Male Not Hispanic/Latino and Black/African American Only, MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other,[1] FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other[1]

Age * Race/Hispanic Recode (4 Levels)
W: Age * Not Hispanic/Latino and White Only, B: Age * Not Hispanic/Latino and Black/African American Only, H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other[1]

Age Squared * Race/Hispanic Recode (4 Levels)
W: Age Squared * Not Hispanic/Latino and White Only, B: Age Squared * Not Hispanic/Latino and Black/African American Only, H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other[1]

Age Cubed * Race/Hispanic Recode (4 Levels)
W: Age Cubed * Not Hispanic/Latino and White Only, B: Age Cubed * Not Hispanic/Latino and Black/African American Only, H: Age Cubed * Hispanic/Latino, O: Age Cubed * Not Hispanic/Latino and Other[1]

Marital Status (3 Levels)
M: Married, WM: Was Married, NM: Never Been Married[1]

Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]

Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

Census Region
N: Northeast, M: Midwest, S: South, W: West[1]

MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Income State Rank
L: Low State Income Rank (lowest tertile), M: Middle State Income Rank (middle tertile), H: High State Income Rank (highest tertile)[1]

Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]

Percentage Non-Hispanic/Latino Black/African American in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

Imputation-Revised Family Received Social Security or Railroad Payments
Y: Yes, N: No[1]

Imputation-Revised Family Received Supplemental Security Income
Y: Yes, N: No[1]

Imputation-Revised Family Received Public Assistance
Y: Yes, N: No[1]

Imputation-Revised Family Received Welfare/Job Placement/Childcare Services
Y: Yes, N: No[1]

Exhibit D.4 Definitions of Levels for Variables (continued)

Imputation-Revised Family Received Income from a Job Y: Yes, N: No[1] **Imputation-Revised Family Received Food Stamps** Y: Yes, N: No[1] **Imputation-Revised Total Family Income > or < $20,000** GT: ≥ $20,000, LT: < $20,000[1] **Intermediate Family Received Social Security or Railroad Payments** Y: Yes, N: No[1] **Intermediate Family Received Supplemental Security Income** Y: Yes, N: No[1] **Intermediate Family Received Public Assistance** Y: Yes, N: No[1] **Intermediate Family Received Welfare/Job Placement/Childcare Services** Y: Yes, N: No[1] **Intermediate Family Received Income from a Job** Y: Yes, N: No[1] **Intermediate Family Received Food Stamps** Y: Yes, N: No[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.53 Dichotomous Income Indicators in Response Propensity Models

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
12-17	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age Cubed * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)

Table D.53 Dichotomous Income Indicators in Response Propensity Models (continued)

Age Group	Variables Included in Response Propensity (Dichotomous Income Indicators)
26-64	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (FT, PT, UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)
65+	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); **Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH)**; **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Marital Status (4 Levels) (M, W, DS)**; **Education Level (4 Levels) (L, HS, SC)**; **Employment Status (FT, PT, UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H, M)**; **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; **Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)**
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)
Number of Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

Table D.54 Dichotomous Income Indicators in Predictive Mean Models: Persons 12 to 17 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; **Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)**
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)

Table D.55 Dichotomous Income Indicators in Predictive Mean Models: Persons 18 to 25 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Number of Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)
Welfare Payments	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y)
Welfare Services	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y)

Table D.56 Dichotomous Income Indicators in Predictive Mean Models: Persons 26 to 64 Years Old (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Number of Welfare Months	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Social Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)
Supplemental Security	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y)
Wages	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y)

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Food Stamps	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y)
Welfare Payments	Age; **Age Squared**; **Age Cubed**; **Gender (M)**; Race/Hispanic Recode (4 Levels) (W, B, H); **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, PT, **UN)**; Census Region (N, M, S); **MSA (**R, **SM)**; Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; **Imputation-Revised Number of Persons Younger than 18 Years Old in Household**; **Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous)**; Intermediate Family Received Social Security or Railroad Payments (Y); **Intermediate Family Received Supplemental Security Income (Y)**
Welfare Services	Age; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; **Age * Gender**; **Age Squared * Gender**; Age * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (**FT, **PT**, **UN)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); **Intermediate Family Received Public Assistance (Y)**

Table D.57 Dichotomous Income Indicators in Predictive Mean Models: Persons 65 Years or Older (continued)

Variable Requiring Imputation	Variables Included in Income Model (Dichotomous Income Indicators)
Number of Welfare Months	**Age**; **Age Squared**; **Age Cubed**; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; **Census Region (**N, M, **S)**; MSA (R, SM); **Income State Rank (L**, **M)**; **Percentage Hispanic/Latino in Segment (**H, **M)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (**H, **M)**; Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); **Intermediate Family Received Social Security or Railroad Payments (Y)**; Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)
Total Income	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Intermediate Family Received Social Security or Railroad Payments (Y); Intermediate Family Received Supplemental Security Income (Y); Intermediate Family Received Public Assistance (Y); Intermediate Family Received Welfare/Job Placement/Childcare Services (Y); Intermediate Family Received Income from a Job (Y); Intermediate Family Received Food Stamps (Y)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.58 Income Finer Categories in Response Propensity Models

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
12-17	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (W**, B, H**)**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, MB, MH**)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); **Imputation-Revised Family Received Food Stamps (Y)**; **Imputation-Revised Total Family Income > or < $20,000 (GT)**
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
26-64	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Marital Status (4 Levels) (M**, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Imputation-Revised Number of Adults in Household**; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

Table D.58 Income Finer Categories in Response Propensity Models (continued)

Age Group	Variables Included in Response Propensity for Income Models (Finer Categorization)
65+	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W, B, H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

Table D.59 Income Finer Categories in Predictive Mean Models

Age Group	Variables Included in Income Models (Finer Categorization)
12-17	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
18-25	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (3 Levels) (M, WM); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)
26-64	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

Table D.59 Income Finer Categories in Predictive Mean Models (continued)

Age Group	Variables Included in Income Models (Finer Categorization)
65+	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); MSA (R, SM); Income State Rank (L, M); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Imputation-Revised Number of Adults in Household; Imputation-Revised Number of Persons Younger than 18 Years Old in Household; Imputation-Revised Number of Persons Older than 64 Years Old in Household (Continuous); Imputation-Revised Family Received Social Security or Railroad Payments (Y); Imputation-Revised Family Received Supplemental Security Income (Y); Imputation-Revised Family Received Public Assistance (Y); Imputation-Revised Family Received Welfare/Job Placement/Childcare Services (Y); Imputation-Revised Family Received Income from a Job (Y); Imputation-Revised Family Received Food Stamps (Y); Imputation-Revised Total Family Income > or < $20,000 (GT)

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.4 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

D.7 Health Insurance Variables

Exhibit D.5 Definitions of Levels for Variables

Gender
M: Male, F: Female[1]

Race/Hispanic Recode (4 Levels)
W: Not Hispanic/Latino and White Only, B: Not Hispanic/Latino and Black/African American Only, H: Hispanic/Latino, O: Not Hispanic/Latino and Other[1]

Age * Race/Hispanic Recode (4 Levels)[2]
W: Age * Not Hispanic/Latino and White Only,[1] B: Age * Not Hispanic/Latino and Black/African American Only, H: Age * Hispanic/Latino, O: Age * Not Hispanic/Latino and Other[1]

Age Squared * Race/Hispanic Recode (4 Levels)[2]
W: Age Squared * Not Hispanic/Latino and White Only,[1] B: Age Squared * Not Hispanic/Latino and Black/African American Only, H: Age Squared * Hispanic/Latino, O: Age Squared * Not Hispanic/Latino and Other[1]

Age Cubed * Race/Hispanic Recode (4 Levels)
W: Age Cubed * Not Hispanic/Latino and White Only, B: Age Cubed * Not Hispanic/Latino and Black/African American Only, H: Age Cubed * Hispanic/Latino, O: Age Cubed * Not Hispanic/Latino and Other[1]

Gender * Race/Hispanic Recode (4 Levels)[2]
MW: Male Not Hispanic/Latino and White Only,[1] MB: Male Not Hispanic/Latino and Black/African American Only, MH: Male Hispanic/Latino, MO: Male Not Hispanic/Latino and Other,[1] FW: Female Not Hispanic/Latino and White Only,[1] FB: Female Not Hispanic/Latino and Black/African American Only,[1] FH: Female Hispanic/Latino,[1] FO: Female Not Hispanic/Latino and Other[1]

Marital Status (2 Levels)
M: Married, NM: Not Married[1]

Marital Status (4 Levels)
M: Married, W: Widowed, DS: Divorced/Separated, NM: Never Been Married[1]

Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]

Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]

MSA
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]

Family Income Recode
L: Income Less than $20,000, M1: Income $20,000-$49,999, M2: Income $50,000-$74,999, H: Income $75,000 or More[1]

Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]

Percentage Non-Hispanic/Latino Black/African American in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[1]

Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10-50)%, L: < 10%[1]

Household Size
L: At Least Four Persons in Household, S: Fewer than Four Persons in Household[1]

Other Family Members In Household
N: No, Y: Yes[1]

Lifetime Military Service
N: No,[1] Y: Yes

Family Wages
N: No,[1] Y: Yes

Family Participation in Government Assistance Programs
N: No,[1] Y: Yes

Family Social Security
N: No,[1] Y: Yes

Exhibit D.5 Definitions of Levels for Variables (continued)

Intermediate MEDICAID/CHIP Coverage N: No,[1] Y: Yes **Intermediate MEDICARE Coverage** N: No,[1] Y: Yes **Intermediate CHAMPUS Coverage** N: No,[1] Y: Yes

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

[2] Under the old method, the reference cell for the race/Hispanic recode was "Not Hispanic/Latino and Other." Under the constituent variables method, the reference cell for the race/Hispanic recode was "Not Hispanic/Latino and White Only." The latter choice is slightly preferable because it facilitates comparisons to the most populous category in the United States. Interactions involving the race/Hispanic recode are similarly handled. See Chapter 9 for details on the processing of the health insurance variables, as well as a fuller description of the differences between the old and constituent variables methods applied to them.

Table D.60 Health Insurance, Constituent Variables Method: Response Propensity Models

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
12-17	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)

Table D.60 Health Insurance, Constituent Variables Method: Response Propensity Models (continued)

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
18-25	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
26-64	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)

Table D.60 Health Insurance, Constituent Variables Method: Response Propensity Models (continued)

Age Group	Set of Variables Used to Determine Nonresponse	Variables Included in Response Propensity Model
	Other Health Insurance[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)
65+	Medicaid/CHIP, Medicare, CHAMPUS, Private Health Insurance	Age; **Age Squared**; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M)
	Other Health Insurance[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 26-64 and 65+ age groups were included in the same response propensity model for other health insurance.

Table D.61 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.62 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); **Lifetime Military Service (Y)**; Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.

Table D.63 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Lifetime Military Service (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance[2]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 18-25 and 26-64 age groups were included in the same predictive mean model for Medicare.

[2] The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

Table D.64 Health Insurance, Constituent Variables Method: Predictive Mean Models, Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Medicaid/CHIP	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Household Size (L); Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)
Medicare	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Percentage Owner Occupied in Segment (H, M); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y)
CHAMPUS	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Lifetime Military Service (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y)
Private Health Insurance	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Marital Status (2 Levels) (M); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L); Percentage Owner Occupied in Segment (H, M); Household Size (L); Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y); Intermediate MEDICAID/CHIP Coverage (Y); Intermediate MEDICARE Coverage (Y); Intermediate CHAMPUS Coverage (Y)
Other Health Insurance[1]	Age; Age Squared; Gender (M); Race/Hispanic Recode (4 Levels) (B, H, O); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (B, H, O); Age Squared * Race/Hispanic Recode (4 Levels) (B, H, O); Gender * Race/Hispanic Recode (4 Levels) (MB, MH, MO); Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); MSA (R, SM); Family Income Recode (L, M1, M2); Percentage Owner Occupied in Segment (H, M); Household Size; Other Family Members In Household (N); Family Wages (Y); Family Participation in Government Assistance Programs (Y); Family Social Security (Y)

NOTE: An asterisk "*" represents an interaction between two variables.

[1] The 26-64 and 65+ age groups were included in the same predictive mean model for other health insurance.

Table D.65 Old Method Health Insurance: Response Propensity Models

Age Group	Variables Included in Response Propensity Model
12-17	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
18-25	**Age**; **Age Squared**; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
26-64	Age; Age Squared; **Age Cubed**; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); **Marital Status (4 Levels) (**M, **W**, **DS)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
65+	**Age**; **Age Squared**; **Age Cubed**; Gender (M); **Race/Hispanic Recode (4 Levels) (**W, B, **H)**; **Age * Gender**; **Age Squared * Gender**; **Age Cubed * Gender**; **Age * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Age Cubed * Race/Hispanic Recode (4 Levels) (W**, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (**MW, MB, **MH)**; Marital Status (4 Levels) (M, W, DS); Education Level (4 Levels) (L, HS, SC); **Employment Status (FT**, **PT**, **UN)**; MSA (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.66 Old Method Health Insurance: Predictive Mean Models, Persons 12 to 17 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.67 Old Method Health Insurance: Predictive Mean Models, Persons 18 to 25 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.68 Old Method Health Insurance: Predictive Mean Models, Persons 26 to 64 Years Old

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: An asterisk "*" represents an interaction between two variables.

Table D.69 Old Method Health Insurance: Predictive Mean Models, Persons 65 Years or Older

Variable Requiring Imputation	Variables Included in Predictive Mean Model
Overall Health Insurance (INSUR3)	Age; Age Squared; Age Cubed; **Gender (M)**; **Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; **Age * Gender**; **Age Squared * Gender**; **Age * Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; **Age Squared * Race/Hispanic Recode (4 Levels) (**W, **B**, **H)**; **Gender * Race/Hispanic Recode (4 Levels) (MW**, **MB**, **MH)**; MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size
Private Health Insurance	Age; Age Squared; Age Cubed; Gender (M); Race/Hispanic Recode (4 Levels) (W, B, H); Age * Gender; Age Squared * Gender; Age * Race/Hispanic Recode (4 Levels) (W, B, H); Age Squared * Race/Hispanic Recode (4 Levels) (W, B, H); Gender * Race/Hispanic Recode (4 Levels) (MW, MB, MH); MSA (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Non-Hispanic/Latino Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Household Size

NOTE: An asterisk "*" represents an interaction between two variables.

D.8 Pair Relationship Variables

Exhibit D.6 Definitions of Levels for Variables

Age Category (Older)
Y: 12-14, T1: 15-17, T2: 18-20, T3: 21-25, A1: 26-34, A2: 35-49, A3: 50+, UK: Unknown[1]
Age Category - Non-Pair
Y: 12-14, T1: 15-17, T2: 18-20, T3: 21-25, A1: 26-34, A2: 35-49, A3: 50+, UK: Unknown[1]
Gender
M: Male,[1] F: Female
Gender (Older)
M: Male,[1] F: Female
Gender (Younger)
M: Male,[1] F: Female
Race/Hispanic Recode (4 Levels)
H: Hispanic/Latino, O: Not Hispanic/Latino and Other, B: Not Hispanic/Latino and Black/African American Only, W: Not Hispanic/Latino and White Only[1]
Race/Hispanic Recode (4 Levels) - Non-Pair
H: Hispanic/Latino, O: Not Hispanic/Latino and Other, B: Not Hispanic/Latino and Black/African American Only, W: Not Hispanic/Latino and White Only[1]
Marital Status (Older) (2 Levels)
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Marital Status (2 Levels) - Non-Pair
MWD: Married, Widowed or Divorced, NM: Never Been Married[1]
Marital Status (Older) (4 Levels)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Marital Status (Younger) (4 Levels)
M: Married, W: Widowed, D: Divorced, NM: Never Been Married[1]
Education Level (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Education Level (Older) (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Education Level (Younger) (4 Levels)
L: Less than High School, HS: High School Graduate, SC: Some College, C: College Graduate[1]
Employment Status
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Employment Status (Older)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Employment Status (Younger)
FT: Full Time, PT: Part Time, UN: Unemployed, OE: Other Employment[1]
Census Region
N: Northeast, M: Midwest, S: South, W: West[1]
Population Density
R: Non-MSA/Rural, SM: Small/Medium MSA, L: Large MSA[1]
Percentage Hispanic/Latino in Segment
H: ≥ 71%, M: [20,71)%, L: < 20%[1]
Percentage Black/African American in Segment
H: ≥ 40%, M: [10,40)%, L: < 10%[1]
Percentage Owner Occupied in Segment
H: ≥ 50%, M: [10,50)%, L: < 10%[1]

NOTE: An asterisk "*" represents an interaction between two variables.

MSA = metropolitan statistical area.

[1] This is the reference level for this variable, against which effects of other factor levels are measured.

Table D.70 Model Summaries (Pair Relationships)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
0 (12-14, 12-14)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); **Race/Hispanic Recode (4 Levels) (**H, **O**, **B)**; **Census Region (N**, M, **S)**; **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, M); **Percentage Black/African American in Segment (H**, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); **Population Density (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); **Population Density (R**, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
1 (12-14, 15-17)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
2 (12,14, 18-25)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
3 (15-17, 15-17)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**; Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
4 (15-17, 18-25)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
5 (18-20, 18-25)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (Older) (T3); Gender (Older) (F); **Gender (Younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (M**, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); **Education Level (Younger) (4 Levels) (L**, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3); Gender (Older) (F); **Gender (Younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (M**, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); **Education Level (Younger) (4 Levels) (L**, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
6 (21-25, 21-25)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); **Gender (Younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (**M, **W**, D**)**; **Education Level (Older) (4 Levels) (**L, **HS**, **SC)**; **Education Level (Younger) (4 Levels) (**L, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
7 (12-14, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); **Race/Hispanic Recode (4 Levels) (**H, **O**, **B)**; Census Region (N, M, S); Population Density (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Black/African American in Segment (**H, **M)**; Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
8 (15-17, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; **Age Category (Older) (A2**, **A3)**; Gender (Older) (F); Gender (Younger) (F); **Race/Hispanic Recode (4 Levels) (H**, **O**, **B)**; **Marital Status (Older) (4 Levels) (M**, **W**, **D)**; **Education Level (Older) (4 Levels) (**L, HS, **SC)**; **Employment Status (Older) (**FT, PT, **UN)**; **Census Region (N**, M, S**)**; **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (Older) (A2**, **A3)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (M**, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
9 (18-20, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (**M, **W**, **D)**; **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); **Employment Status (Older) (**FT, **PT**, **UN)**; Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); **Percentage Hispanic/Latino in Segment (H**, **M)**; Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); **Education Level (Younger) (4 Levels) (L**, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); **Marital Status (Younger) (4 Levels) (**M, **W**, **D)**; Education Level (Older) (4 Levels) (L, HS, SC); **Education Level (Younger) (4 Levels) (L**, **HS**, **SC)**; Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.70 Model Summaries (Pair Relationships) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
10 (21+, 26+)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); Marital Status (Younger) (4 Levels) (M, W, D); Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Age Category (Older) (A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (4 Levels) (M, W, D)**; **Marital Status (Younger) (4 Levels) (M, W, D)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (Older) (A2, A3)**; Gender (Older) (F); **Gender (Younger) (F)**; Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (4 Levels) (M, W, D); **Marital Status (Younger) (4 Levels) (M, W, D)**; Education Level (Older) (4 Levels) (L, HS, SC); Education Level (Younger) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Employment Status (Younger) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.71 Model Summaries (Multiplicities)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Child Focus	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); **Age Category (Older) (T2**, **T3**, **A1**, **A2)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category (Older) (T2**, **T3**, **A1**, **A2)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
Parent-Child (12-20) Parent Focus	No Model Used; No Nonrespondents	No Model Used; No Nonrespondents	No Model Used; No Nonrespondents
Sibling (12-14) Sibling (15-17) Older Sibling Focus	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.71 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17) Younger Sibling Focus	**Household Size (Continuous)**; Household Size (5 Categories); Gender (Older) (F); Gender (Younger) (F); **Race/Hispanic Recode (4 Levels) (**H, **O**, **B)**; **Census Region (**N, **M**, S**)**; **Population Density (R**, **SM)**; **Percentage Hispanic/Latino in Segment (H**, **M)**; **Percentage Black/African American in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (**H, **M)**	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Sibling (12-17) Sibling (18-25) Older Sibling Focus	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T2); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.71 Model Summaries (Multiplicities) (continued)

Pair Domain	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25) Younger Sibling Focus	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T2); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T2); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T2); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Child Focus, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Parent-Child (12-20) Child Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Parent Focus, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
Parent-Child (12-20) Parent Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Sibling (12-14) Sibling (15-17), Older Sibling Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Sibling (12-17) Sibling (18-25), Older Sibling Focus, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); **Race/Hispanic Recode (4 Levels) (**H, O, **B)**; Census Region (N, M, S); Population Density (R, SM); **Percentage Hispanic/Latino in Segment (**H, **M)**; **Percentage Black/African American in Segment (H**, **M)**; **Percentage Owner Occupied in Segment (H**, **M)**	Age Category (Older) (Y); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Census Region (N, M, S); Population Density (R, SM); **Percentage Hispanic/Latino in Segment (**H, **M)**; Percentage Black/African American in Segment (H, M); **Percentage Owner Occupied in Segment (H**, **M)**; Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
Spouse-Spouse, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); **Marital Status (Older) (2 Levels) (MWD)**; Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse with Children, Both Pair Members Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.72 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse with Children, at Least One Pair Member Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category (Older) (T1, T2, T3, A1, A2, A3); Gender (Older) (F); Gender (Younger) (F); Race/Hispanic Recode (4 Levels) (H, O, B); Marital Status (Older) (2 Levels) (MWD); Education Level (Older) (4 Levels) (L, HS, SC); Employment Status (Older) (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Child Focus, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Parent-Child (12-20) Child Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Parent-Child (12-20) Parent Focus, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; Household Size (5 Categories); Age Category - Non-Pair (Y); Gender (F); **Race/Hispanic Recode (4 Levels) - Non-Pair (**H, O, **B)**; Census Region (N, M, S); **Population Density (**R, **SM)**; Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
Parent-Child (12-20) Parent Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Age Category - Non-Pair (T3**, **A1**, **A2**, **A3)**; Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (2 Levels) - Non-Pair (MWD)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Sibling (12-14) Sibling (15-17), Older Sibling Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+
Sibling (12-17) Sibling (18-25), Older Sibling Focus, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; Number in Household Aged 35-49; Number in Household Aged 50+

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse, Younger than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (Y); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); Number in Household Aged 0-11; Number in Household Aged 12-17; Number in Household Aged 18-25; Number in Household Aged 26-34; **Number in Household Aged 35-49**; **Number in Household Aged 50+**
Spouse-Spouse, Older than 18	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (2 Levels) - Non-Pair (MWD)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); **Marital Status (2 Levels) - Non-Pair (MWD)**; Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

Table D.73 Model Summaries (Household-Level Person Counts of Pair Domains when Respondent Is Not in a Responding Pair) (continued)

Model Group	Variables Included in Response Propensity Model	Variables Included in Predictive Mean Model	
		Including Household Size	Not Including Household Size
Spouse-Spouse with Children	Household Size (Continuous); **Household Size (5 Categories)**; Age Category - Non-Pair (T1, T2, T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	**Household Size (Continuous)**; **Household Size (5 Categories)**; **Age Category - Non-Pair (**T1, T2, T3, A1, A2, **A3)**; Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M)	Age Category - Non-Pair (T1, T2, T3, A1, A2, A3); Gender (F); Race/Hispanic Recode (4 Levels) - Non-Pair (H, O, B); Marital Status (2 Levels) - Non-Pair (MWD); Education Level (4 Levels) (L, HS, SC); Employment Status (FT, PT, UN); Census Region (N, M, S); Population Density (R, SM); Percentage Hispanic/Latino in Segment (H, M); Percentage Black/African American in Segment (H, M); Percentage Owner Occupied in Segment (H, M); **Number in Household Aged 0-11**; **Number in Household Aged 12-17**; **Number in Household Aged 18-25**; **Number in Household Aged 26-34**; **Number in Household Aged 35-49**; **Number in Household Aged 50+**

NOTE: Boldface of variables and levels indicates that they were dropped from the model.
NOTE: See Exhibit D.6 for definitions of levels for variables.
NOTE: An asterisk "*" represents an interaction between two variables.

Appendix E: Hot-Deck Procedure Summaries

Appendix E: Hot-Deck Procedure Summaries

E.1 Introduction

For the majority of variables that had missing values imputed in the 2012 National Survey on Drug Use and Health (NSDUH), the imputation method used was predictive mean neighborhood (PMN). This appendix summarizes the predictive mean vectors and the constraints applied during the PMN hot-deck step.[1] It is organized by groups of variables requiring imputation: demographics, household composition (roster), lifetime use of drugs, recency and frequency of drug use, age at first drug use, income, health insurance, and pair. There are three types of tables associated with each variable or set of variables imputed using the PMN method: (1) Logical Constraints; (2) Likeness Constraints; and (3) Constraints and Portion of the Predictive Mean Vector.

For variables that do not apply any logical constraints in the PMN process, the "Logical Constraints" table is not applicable. The "Constraints and Portion of the Predictive Mean Vector" table specifies the following for each missingness pattern:

1. the number of item nonrespondents exhibiting the pattern ("Total Number of Cases");
2. the set of logical constraints applied to the potential donors ("Logical Constraints");
3. the elements of the predictive mean vector ("Predictive Mean Vector") used to calculate the Mahalanobis distance from recipient to potential donor, as well as to restrict the donor set via the delta constraint; and
4. the set of likeness constraints utilized in each try and the number of item nonrespondents who found donors on each try by different age groups ("Likeness Constraints: Number of Cases").[2]

In the tables that follow, the phrase "Donor's predicted means each must be within x percent of recipient's predicted means" appears in each of the multivariate imputation tables, and the phrase "Donor's predicted mean must be within x percent of recipient's predicted mean" appears in each of the univariate imputation tables. In either case, it represents one of the likeness constraints, known as the "delta constraint," and also defines the neighborhood. When this constraint is loosened, the neighborhood of potential donors is abandoned, and the candidate with the predicted mean closest to the recipient's (subject to the constraints that are still on the pool of donors) is chosen as the donor.

Although statistical imputation of the drug use or income variables could not have proceeded separately within each State because of insufficient pools of donors, the PMN procedure does incorporate information about the State of residence of each respondent. For the drug use variables, in the hot-deck step of PMN, respondents were separated into three State

[1] See Chapter 3 for details on PMN imputation.

[2] If a cell contains the text "No Cases" within "Likeness Constraints: Number of Cases," this means that there were no cases requiring imputation in that category.

usage-level categories for each drug, depending on the response variable of interest. Respondents from States with high usage of a given drug were placed in one category, respondents from medium-usage States were placed in another category, and respondents from low-usage States were placed in a third category. For the income variables, the States were separated into three income groups, depending upon the proportion of families within those States with incomes greater than or equal to $20,000. As with the drug use variables, respondents from high-income States (by this measure) were placed in one category, respondents from medium-income States were placed in another category, and respondents from low-income States were placed in a third category. In the tables that follow, this variable is identified as the "State rank" for the drug use and income variables. It was used as a likeness constraint, where the set of eligible donors for each recipient was restricted so that donors and recipients were both from States with the same State rank.

E.2 Demographics

Tables E.1 through E.11 present information on the imputation procedures for the core demographic variables: marital status, race, Hispanic/Latino origin, Hispanic/Latino group, and education level. Tables E.12 through E.19 present information for the noncore demographic variables: employment status, indicator of birth in the United States, and immigrant age of entry into the United States. In several instances, variable names are used without description for the purposes of brevity (see Chapter 4 for details). The segment-level variable, SEGID (Segment ID), was used only in the likeness constraints for demographic imputation.

As described in the 2012 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013) within each State, State sampling (SS) regions were formed, which were further partitioned into clusters of adjacent blocks called "segments." The segment ID number was a two-letter State abbreviation followed by a two-digit SS region and a two-digit segment identifier, which uniquely identifies each segment. Although the segment identifier was not used as a covariate because of the large number of levels, it was used as a constraint in the hot-deck step of the PMN procedure for race, Hispanicity, education, and employment status, as noted in Chapters 4 and 5. For more information regarding segments, see the 2012 NSDUH sample design report (Morton et al., 2013).

E.2.1 Marital Status Variables

Table E.1 Likeness Constraints for Marital Status

Constraint #	Likeness Constraint
LikC1	Donor's age must be within 3 years of recipient's age
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.2 Constraints and Portion of the Predictive Mean Vector for Marital Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	8	None	1. M1 2. M2 3. M3	1,2: 1	1,2: 2	1,2: 5

[1] The predictive mean vector components are defined by the following:
1. M1 = P(Married)
2. M2 = P(Widowed)
3. M3 = P(Divorced or separated)

E.2.2 Race Variables

Table E.3 Logical Constraints for Race

Constraint #	Logical Constraint
LogC1	Donor must be Asian, in part or in full
LogC2	Donor must be more than one race
LogC3	Donor must not be white only
LogC4	Donor must be white only or white and American Indian/Alaska Native only
LogC5	Donor must not be American Indian/Alaska Native, in part or in full

Table E.4 Likeness Constraints for Race

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC3	If recipient was Hispanic/Latino nonspecific, donor must be of Hispanic/Latino origin
LikC4	If recipient selected one or more Hispanic/Latino categories, including Mexican, Puerto Rican, Central or South American, Cuban, Dominican, and Spaniard, then donor's Hispanic/Latino group value must be equal to one of the Hispanic/Latino groups mentioned by recipient
LikC5	Donor must be Mexican (Hispanic/Latino or non-Hispanic/Latino)
LikC6	Donor must be Cuban (Hispanic/Latino or non-Hispanic/Latino)
LikC7	Donor must be Central or South American (Hispanic/Latino or non-Hispanic/Latino)
LikC8	Donor must be Dominican (Hispanic/Latino or non-Hispanic/Latino)
LikC9	Donor must be Spanish (Hispanic/Latino or non-Hispanic/Latino)
LikC10	Donor must be Puerto Rican (Hispanic/Latino or non-Hispanic/Latino)

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	2,372	None	1. R1 2. R2 3. R3 4. R4	1-4: 367 2-4: 403 3,4: 94	1-4: 154 2-4: 441 3,4: 252	1-4: 30 2-4: 217 3,4: 414
2	Known to be Asian	7	1	1. R1 2. R2 3. R3 4. R4	1-4: 1 2-4: 0 3,4: 1	1-4: 0 2-4: 0 3,4: 2	1-4: 0 2-4: 2 3,4: 1
3	Known to be multiple race, but no other information	7	2	1. R1 2. R2 3. R3 4. R4	1-4: 0 2-4: 4	1-4: 0 2-4: 0 3,4: 1	1-4: 0 2-4: 0 3,4: 2
4	Known to be nonwhite, but no other information	8	3	1. R2/(1-R1) 2. R3/(1-R1) 3. R4/(1-R1)	1-4: 1 2-4: 2 3,4: 2	No Cases	1-4: 0 2-4: 1 3,4: 2
5	Known to be white or both white and American Indian/Alaska Native	0	4	1. R1/(1-R2-R3-R4)	No Cases	No Cases	No Cases
6	Known not to be American Indian/ Alaska Native, in part or in full	0	5	1. R1/(1-R3) 2. R2/(1-R3) 3. R4/(1-R3)	No Cases	No Cases	No Cases
7	Known to be non-Hispanic Mexican	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
8	Known to be non-Hispanic Cuban	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases

Table E.5 Constraints and Portion of the Predictive Mean Vector for Race (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
9	Known to be non-Hispanic Central or South American	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
10	Known to be non-Hispanic Dominican	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
11	Known to be non-Hispanic Spanish	0	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	No Cases
12	Known to be non-Hispanic Puerto Rican	1	None	1. R1 2. R2 3. R3 4. R4	No Cases	No Cases	1-4,10: 0 2-4,10: 0 3,4,10: 1

[1] The predictive mean vector components are defined by the following:
1. R1 = P(White Only)
2. R2 = P(Black Only)
3. R3 = P(American Indian/Alaska Native Only)
4. R4 = P(Asian Only)

E.2.3 Hispanic/Latino Origin Variables

Table E.6 Likeness Constraints for Hispanic/Latino Origin

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.7 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Origin

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	162	None	1. H1	1,2: 101 2: 38	1,2: 3 2: 6	1,2: 0 2: 14

[1] The predictive mean vector components are defined by the following:
1. H1 = P(Hispanic/Latino origin)

E.2.4 Hispanic/Latino Group Variables

Table E.8 Likeness Constraints for Hispanic/Latino Group

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient had 8 ≤ EDHOGRP ≤ 21, donor's values for IRRACWH, IRRACEBK, IRRACENA, IRRACEAS, IRRACENH, and IRRACEPI must indicate a subset of the racial categories mentioned by recipient

Table E.9 Constraints and Portion of the Predictive Mean Vector for Hispanic/Latino Group

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases[2]
1	Completely missing	68	None	1. H1 2. H2 3. H3	1-3: 14 2,3: 42 3: 12

[1] The predictive mean vector components are defined by the following:
1. H1 = P(Mexican)
2. H2 = P(Puerto Rican)
3. H3 = P(Central or South American)

[2] The hot-deck program for the Hispanic/Latino group is not separated into age groups.

E.2.5 Education Variables

Table E.10 Likeness Constraints for Education Level

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Age of donor = Age of recipient

Table E.11 Constraints and Portion of the Predictive Mean Vector for Education Level

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector, by Age Group[1]		Likeness Constraints: Number of Cases, by Age Group		
				12-17	18+	12-17	18-25	26+
1	Missing	9	None	1. E1 2. E2 3. E3 4. E4	1. E1 2. E2 3. E3	1-3: 1 2,3: 3	1-3: 1 2,3: 3	1,2: 0 2: 1

[1] The predictive mean vector components are defined by the following:
12-17 Age Group
1. E1 = P(Less than elementary school)
2. E2 = P(Elementary school)
3. E3 = P(Middle school)
4. E4 = P(Some high school)

18+ Age Group
1. E1 = P(Less than high school)
2. E2 = P(High school)
3. E3 = P(Some college)

E.2.6 Employment Variables

Table E.12 Logical Constraints for Employment Status

Constraint #	Logical Constraint
LogC1	Donor must be employed

Table E.13 Likeness Constraints for Employment Status

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	Donor's age must be within 5 years of recipient's age

Table E.14 Constraints and Portion of the Predictive Mean Vector for Employment Status

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	17	None	1. E1 2. E2 3. E3	1-3: 0 2,3: 2 3: 1	1-3: 1 2,3: 6	1-3: 0 2,3: 4 3: 3
2	Known to be employed; part-time vs. full-time status unknown	28	1	1. E1/ (E1+E2)	1-3: 0 2,3: 2	1-3: 1 2,3: 14	1-3: 0 2,3: 11

[1] The predictive mean vector components are defined by the following:
1. E1 = P(Employed full time)
2. E2 = P(Employed part time)
3. E3 = P(Unemployed)

E.2.7 Immigrant Variables

Table E.15 Likeness Constraints for Indicator of Birth in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.16 Constraints and Portion of the Predictive Mean Vector for Indicator of Birth in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Completely missing	16	None	1. B1	1,2: 3 2: 6	1,2: 1 2: 3	1,2: 0 2: 3

[1] The predictive mean vector components are defined by the following:
1. B1 = P(Born in the United States)

Table E.17 Logical Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LogC1	Donor's age of entry must be less than recipient's current age
LogC2	Difference between recipient's current age and donor's age of entry must be less than or equal to 1 if recipient lived in the United States less than a year, or difference must be greater than 1 if recipient lived in the United States more than a year

Table E.18 Likeness Constraints for Age of Entry in the United States

Constraint #	Likeness Constraint
LikC1	Segment of donor = Segment of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean

Table E.19 Constraints and Portion of the Predictive Mean Vector for Age of Entry in the United States

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases[2]
1	Completely missing	15	1,2	1. PrAgeEntry	1,2: 1 2: 14

[1] The predictive mean vector components are defined by the following:
1. PrAgeEntry = Predicted age of entry

[2] The hot-deck program for immigrant age of entry is not separated into age groups.

E.3 Drug Variables

Tables E.20 through E.75 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures to variables for lifetime drug use, recency and frequency of use, and age at first use.

E.3.1 Lifetime Drug Use

There were a large number of missingness patterns for lifetime drug use. The usable case rule (Section 2.2.3) required a certain minimum response pattern, but other than that any combination of the lifetime drug variables may have been missing for a given respondent record.

The probabilities associated with the 14 lifetime drug variables listed in Table E.20 formed the full predictive mean vector. Only the probabilities associated with the lifetime drug variables for which values were missing were used in the predictive mean vector for each item nonrespondent. The predicted mean for cigarette lifetime use was never calculated because no respondents were missing the lifetime variable for cigarette use.[3]

Note that if only a child lifetime drug variable was missing, the predicted mean associated with the parent drug was used. No predicted means were calculated for child drugs. For example, if the only missing lifetime drug use variable was the one for methamphetamine, the predictive mean "vector" would be a scalar: the predicted mean associated with lifetime use of any stimulant.[4]

[3] The lifetime variable for cigarette use was never missing because the usable case rule required a nonmissing response to the cigarette gate question.

[4] Crack is an exception to this rule. Crack is considered a child drug of cocaine, but it is unique among child drugs in that it is covered by a separate module in the questionnaire. Those denying cocaine lifetime use are automatically assigned a negative response for crack lifetime use, and those with a missing response for cocaine lifetime use are automatically assigned a missing response for crack lifetime use.

Table E.20 Elements of Full Predictive Mean Vector for Lifetime Drug Use

Lifetime Drug Use	Predicted Mean
Heroin Lifetime	P(Lifetime User)
Crack Lifetime	P(Lifetime User | Lifetime User of Cocaine), if cocaine lifetime use is known P(Lifetime User of Cocaine)* P(Lifetime User | Lifetime User of Cocaine), if cocaine lifetime use is unknown
Cocaine Lifetime	P(Lifetime User)
Sedatives Lifetime	P(Lifetime User)
Stimulants/Methamphetamine Lifetime	P(Lifetime User)
Tranquilizers Lifetime	P(Lifetime User)
Pain Relievers/OxyContin Lifetime	P(Lifetime User)
Hallucinogens/LSD/PCP/Ecstasy Lifetime	P(Lifetime User)
Marijuana Lifetime	P(Lifetime User)
Inhalants Lifetime	P(Lifetime User)
Alcohol Lifetime	P(Lifetime User)
Pipes Lifetime	P(Lifetime User)
Snuff/Chewing Tobacco Lifetime	P(Lifetime User)
Cigars Lifetime	P(Lifetime User)

Table E.21 Logical Constraints for Lifetime

Constraint #	Logical Constraint
LogC1	Donor must be a lifetime user of pain relievers if recipient is known to have used pain relievers but missing both indicators for OxyContin and "other" pain reliever

Table E.22 Likeness Constraints for Lifetime

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Lifetime use of donor = Lifetime use of recipient, for each nonmissing lifetime indicator
LikC4	If recipient was missing the lifetime indicator(s) for any member of a family of drugs,[1] donor's lifetime indicator(s) agreed with recipient's nonmissing lifetime indicator(s) within that family

[1] The smokeless tobacco family includes chewing tobacco and snuff. The hallucinogens family includes LSD, PCP, Ecstasy, and other hallucinogens. The pain relievers family includes OxyContin and other pain relievers. The stimulants family includes methamphetamine and other stimulants. The cocaine family includes crack.

Table E.23 Constraints and Portion of the Predictive Mean Vector for Lifetime

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing one or more of the lifetime drug use variables	890	1	1. LF for gate question associated with each missing variable	1-3: 492 1,2,4: 82 1,4: 30	1-3: 90 1,2,4: 42 1,4: 21	1-3: 71 1,2,4: 48 1,4: 14

[1] The predictive mean vector components are defined by the following:
1. LF = P(Lifetime user of x), where x is the drug associated with the missing gate question

Table E.24 Likeness Constraints for Cigarette Ever Daily Used

Constraint #	Likeness Constraint
LikC1	Age of donor = Age of recipient
LikC2	State rank of donor = State rank of recipient
LikC3	If recipient was a past year cigarette ever daily user, then donor was a past year cigarette ever daily user; if recipient was a past 3 years but not past year cigarette ever daily user, then donor was the same; if recipient was a lifetime but not past 3 years cigarette ever daily user, then donor was the same
LikC4	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC5	Age of donor must be less than or equal to age of recipient

Table E.25 Constraints and Portion of the Predictive Mean Vector for Cigarette Ever Daily Used

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing CIGDLYMO	14	None	1. CIG	1-4: 1	1-4: 7	1-4: 3 1,3,4: 1 1,3: 2

[1] The predictive mean vector components are defined by the following:
1. CIG = P(Daily use at some point in lifetime | respondent was a lifetime user but not a current daily user)

E.3.2 Drug Recency and Frequency

Many tables in this section abbreviate certain words. "Recency" is an abbreviation for "recency of use," "frequency" or "Freq." is an abbreviation for "frequency of use," and "30-day binge drink" or "DR5DAY" is an abbreviation for the "number of days in the past 30 days when the respondent consumed five or more alcoholic drinks."

An empty cell within "Missingness Pattern" indicates that all information was available. A entry of "Missing" in a cell indicates that all information was missing. Other entries for missingness patterns provide the available information, indicating that this information was partially missing. For example, an entry of "Past year" indicates that the individual used the drug of interest in the past year, but the specific recency of past month versus past year not past month must be imputed. However, if the entry is shown in parentheses, then the specific recency is known and no imputation is required. For example, an entry of "(Past year)" indicates that the individual used the drug of interest in the past year and the specific recency of past month versus past year not past month is already known and does not require imputation.

Table E.26 Logical Constraints for Cigarette Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used cigarettes within the past 3 years (a cigarette recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month cigarette user (cigarette recency ≠ 1)
LogC3	Donor cannot be a past year cigarette user (cigarette recency ≠ 1 or 2)
LogC4	Donor must be a past year cigarette user (cigarette recency = 1 or 2)
LogC5	Donor must be a past month cigarette user (cigarette recency = 1)
LogC6	If recipient was never a daily user of cigarettes (CIGDLYMO = 2), donor's 30-day cigarette frequency cannot equal 30
LogC7	If recipient's age at first use equals his or her current age, donor's 30-day cigarette frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first cigarette use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC8	Donor must be a past year but not past month cigarette user or a past 3 years but not past year cigarette user (cigarette recency = 2 or 3)

Table E.27 Likeness Constraints for Cigarette Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.28 Constraints and Portion of the Predictive Mean Vector for Cigarette Recency and Frequency

#	Missingness Pattern[1]	Total Number of Cases	Logical Constraints	Predictive Mean Vector[2]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	5	4,6	1. R1/(R1+R2) 2. (R1*D)/(R1+R2) 3. R1*(1-D)*PM/(R1+R2)	1,2: 1 1: 3	1,2: 1	No Cases
2	Recency and 30-day frequency missing (lifetime use known)	7	6	1. R1 2. R2 3. R3 4. R1*D 5. R1*(1-D)*PM	1,2: 1 1: 1	1,2: 0 1: 4	1,2: 0 1: 1
3	Past month, 30-day frequency missing	26	5-7	1. D 2. PM	1,2: 5 1: 4	1,2: 14	1,2: 3
4	Not past year	199	3,6	1. R3/(R3+R4)	1,2: 114 1: 2	1,2: 82	1,2: 1
5	Not past month	80	2,6	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 58 1: 2	1,2: 13	1,2: 7
6	Past year but not past month, or past 3 years but not past year	133	6,8	1. R2/(R2+R3)	1,2: 92	1,2: 41	No Cases
7	Past 3 years, 30-day frequency missing	2	1,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*D)/(R1+R2+R3) 4. R1*(1-D)*PM/(R1+R2+R3)	1,2: 0 1: 2	No Cases	No Cases

[1] The response to CIGDLYMO, the edited response to the "ever daily used" question, technically could be used to subdivide each of the first three missingness patterns into two: one for respondents with CIGDLYMO = 2, and the other for respondents with CIGDLYMO ≠ 2. This was not done, because the benefit derived from this change would likely be insignificant. Respondents with CIGDLYMO = 2 technically have zero probability of being a daily user, so the predictive mean vectors could be simplified by setting D = 0. For example, the predictive mean vector for respondents in Missingness Pattern 2 with CIGDLYMO = 2 might look like this: (1) R1; (2) R2; (3) R3; (4) R1*PM.

[2] The predictive mean vector components are defined by the following:
1. R1 = P(Past month cigarette use | lifetime cigarette use)
2. R2 = P(Past year but not past month cigarette use | lifetime cigarette use)
3. R3 = P(Past 3 years but not past year cigarette use | lifetime cigarette use)
4. R4 = P(Lifetime but not past 3 years cigarette use | lifetime cigarette use)
5. D = P(Daily cigarette use | past month cigarette use)
6. PM = P(Cigarette use on a given day in the past month | past month use)

Table E.29 Logical Constraints for Cigar Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used cigars within the past 3 years (a cigar recency category of 1, 2, or 3)
LogC2	Donor cannot be a past month cigar user (cigar recency ≠ 1)
LogC3	Donor must be a past year cigar user (cigar recency = 1 or 2)
LogC4	Donor must be a past month cigar user (cigar recency = 1)
LogC5	If recipient was never a daily user of cigarettes (CIGDLYMO = 2), donor's 30-day cigarette frequency cannot equal 30
LogC6	If recipient's age at first use equals his or her current age, donor's 30-day cigar frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first cigar use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC7	Donor must be a past year but not past month cigar user or a past 3 years but not past year cigar user (cigar recency = 2 or 3)

Table E.30 Likeness Constraints for Cigar Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.31 Constraints and Portion of the Predictive Mean Vector for Cigar Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Past year, 30-day frequency missing	7	3	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1,2: 3 1: 2	1,2: 2	No Cases
2	Recency and 30-day frequency missing (lifetime use known)	10	None	1. R1 2. R2 3. R3 4. R1*PM	1,2: 0 1: 3	1,2: 3 1: 1	1,2: 0 1: 3
3	Past month, 30-day frequency missing	10	4,6	1. PM	1,2: 3	1,2: 5	1,2: 1 1: 1
4	Not past year	122	3	1. R3/(R3+R4)	1,2: 43 1: 4	1,2: 60	1,2: 15
5	Not past month	80	2	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1,2: 30 1: 11	1,2: 31	1,2: 7 1: 1
6	Past year but not past month, or past 3 years but not past year	143	7	1. R2/(R2+R3)	1,2: 79 1: 1	1,2: 55	1,2: 8
7	Past 3 years, 30-day frequency missing	2	1	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. (R1*PM)/(R1+R2+R3)	1,2: 0 1: 1	1,2: 1	No Cases

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month cigar use | lifetime cigar use)
2. R2 = P(Past year but not past month cigar use | lifetime cigar use)
3. R3 = P(Past 3 years but not past year cigar use | lifetime cigar use)
4. R4 = P(Lifetime but not past 3 years cigar use | lifetime cigar use)
5. PM = P(Cigar use on a given day in the past month | past month cigar use)

Table E.32 Logical Constraints for Smokeless Tobacco Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must have used chewing tobacco within the past 3 years (a chew recency category of 1, 2, or 3)
LogC2	Donor must have used snuff within the past 3 years (a chew recency category of 1, 2, or 3)
LogC3	Donor must be a lifetime user of chewing tobacco
LogC4	Donor must be a lifetime user of snuff
LogC5	If recipient's age at first chewing tobacco use equals his or her current age, donor's 30-day chewing tobacco frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first chewing tobacco use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC6	If recipient's age at first snuff use equals his or her current age, donor's 30-day snuff frequency (1) cannot be greater than the number of days between recipient's interview date and his or her date of first snuff use (inclusive) and (2) cannot be greater than the number of days between recipient's interview date and his or her birthday (inclusive)
LogC7	Donor must be a past month chewing tobacco user (chew recency = 1)
LogC8	Donor must be a past month snuff user (snuff recency = 1)
LogC9	Donor's snuff recency must equal recipient's snuff recency
LogC10	Donor's chewing tobacco recency must equal recipient's chewing tobacco recency
LogC11	Donor must have used chewing tobacco within the past year (snuff recency = 1 or 2)
LogC12	Donor must have used snuff within the past year (chew recency = 1 or 2)
LogC13	Donor must be a past 3 years but not past year or lifetime but not past 3 years chewing tobacco user (chew recency = 3 or 4)
LogC14	Donor must be a past 3 years but not past year or lifetime but not past 3 years snuff user (snuff recency = 3 or 4)
LogC15	Donor must be a past year but not past month, past 3 years but not past year, or lifetime but not past 3 years chewing tobacco user (chew recency = 2, 3, or 4)
LogC16	Donor must be a past year but not past month, past 3 years but not past year, or lifetime but not past 3 years snuff user (snuff recency = 2, 3, or 4)
LogC17	Donor must be a past year but not past month or a past 3 years but not past year chewing tobacco user (chew recency = 2 or 3)
LogC18	Donor must be a past year but not past month or a past 3 years but not past year snuff user (snuff recency = 2 or 3)

Table E.33 Likeness Constraints for Smokeless Tobacco Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's recencies for chewing tobacco and snuff agree with recipient's recencies for chewing tobacco and snuff (when nonmissing)
LikC4	Donor's lifetime use statuses for chewing tobacco and snuff agree with recipient's lifetime use statuses for chewing tobacco and snuff (when nonmissing)

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
1	(Past month)	(Past month)	Missing	Missing	0	5-8	1. DC 2. PMC 3. DS 4. PMS	No Cases	No Cases	No Cases
2	(Past month)		Missing		1	5,7	1. DC 2. PMC	No Cases	1-3: 0 1,2,4: 0 1,4: 1	No Cases
3		(Past month)		Missing	1	6,8	1. DS 2. PMS	1-3: 0 1,2,4: 0 1,4: 1	No Cases	No Cases
4		Missing (lifetime use known)		Missing	7	4,6	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	1-3: 0 1,2,4: 0 1,4: 3	1-3: 0 1,2,4: 0 1,4: 4	No Cases
5	(Past month)	Missing (lifetime use known)	Missing	Missing	0	4-7	1. R1 2. R2 3. R3 4. DC 5. PMC 6. RS1*DS 7. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
6	Missing (lifetime use known)		Missing		2	3,5	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	1-3: 0 1,2,4: 0 1,4: 1	1-3: 0 1,2,4: 0 1,4: 1	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
7	Missing (lifetime use known)	(Past month)	Missing	Missing	1	3,5,6,8	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. DS 7. PMS	No Cases	1-3: 0 1,2,4: 0 1,4: 1	No Cases
8		Past year		Missing	5	6,12	1. R1/(R1+R2) 2. RS1*DS/ (RS1+RS2) 3. RS1*(1-DS)*PMS/ (RS1+RS2)	1-3: 0 1,2,4: 0 1,4: 3	1-3: 0 1,2,4: 0 1,4: 2	No Cases
9	Past year		Missing		5	5,11	1. R1/(R1+R2) 2. RC1*DC/ (RC1+RC2) 3. RC1*(1-DC)*PMC/ (RC1+RC2)	1-3: 0 1,2,4: 0 1,4: 4	1-3: 0 1,2,4: 0 1,4: 1	No Cases
10	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	3-6	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC 6. RS1*DS 7. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
11	Not past year				47	13	1. R3/(R3+R4)	1-3: 6 1,2,4: 4 1,4: 9	1-3: 23 1,2,4: 0 1,4: 3	1-3: 2
12		Not past year			75	14	1. R3/(R3+R4)	1-3: 18 1,2,4: 1 1,4: 11	1-3: 39 1,2,4: 3	1-3: 3

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
13	Not past year	Not past year			17	13,14	1. R3/(R3+R4)	1-3: 6 1,2,4: 0 1,4: 1	1-3: 8 1,2,4: 0 1,4: 1	1-3: 1
14	Not past month				26	15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 8 1,2,4: 0 1,4: 4	1-3: 4 1,2,4: 2 1,4: 2	1-3: 0 1,2,4: 0 1,4: 6
15		Not past month			41	16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 14 1,2,4: 1 1,4: 2	1-3: 12 1,2,4: 0 1,4: 5	1-3: 0 1,2,4: 0 1,4: 7
16	Not past month	Not past month			2	15,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 1	1-3: 1	No Cases
17	Not past month	(Past month)		Missing	0	6,8,15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DS 4. PMS	No Cases	No Cases	No Cases
18	(Past month)	Not past month	Missing		0	5,7,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4) 3. DC 4. PMC	No Cases	No Cases	No Cases
19	Not past month	Missing (lifetime use known)		Missing	0	4,6,15	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
20	Missing (lifetime use known)	Not past month	Missing		0	3,5,16	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
21	Not past month	Not past year			0	14,15	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases
22	Not past year	Not past month			0	13,16	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases
23	Not past year	Missing (lifetime use known)		Missing	0	4,6,13	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
24	Missing (lifetime use known)	Not past year	Missing		0	3,5,14	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
25	Past year	Past year	Missing	Missing	1	5,6,11,12	1. R1/(R1+R2) 2. RC1*DC/ (RC1+RC2) 3. RC1*(1-DC)*PMC/ (RC1+RC2) 4. RS1*DS/ (RS1+RS2) 5. RS1*(1-DS)*PMS/ (RS1+RS2)	No Cases	No Cases	1-3: 0 1,2,4: 0 1,4: 1

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
26		Past 3 years		Missing	0	2,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RS1*DS/(RS1+RS2+RS3) 4. RS1*(1-DS)*PMS/(RS1+RS2+RS3)	No Cases	No Cases	No Cases
27	Past 3 years		Missing		0	1,5	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/(RC1+RC2+RC3) 4. RC1*(1-DC)*PMC/(RC1+RC2+RC3)	No Cases	No Cases	No Cases
28	Past year but not past month, or past 3 years but not past year				34	17	1. R2/(R2+R3)	1-3: 13 1,2,4: 1 1,4: 2	1-3: 15 1,2,4: 2	1-3: 0 1,2,4: 0 1,4: 1
29		Past year but not past month, or past 3 years but not past year			59	18	1. R2/(R2+R3)	1-3: 35 1,2,4: 0 1,4: 2	1-3: 19 1,2,4: 0 1,4: 1	1-3: 1 1,2,4: 0 1,4: 1
30	Past year but not past month, or past 3 years but not past year	Past year but not past month, or past 3 years but not past year			9	17,18	1. R2/(R2+R3)	1-3: 5	1-3: 4	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
31	Past year but not past month, or past 3 years but not past year	(Past month)		Missing	0	6,8,17	1. R2/(R2+R3) 2. DS 3. PMS	No Cases	No Cases	No Cases
32	(Past month)	Past year but not past month, or past 3 years but not past year	Missing		0	5,17,18	1. R2/(R2+R3) 2. DC 3. PMC	No Cases	No Cases	No Cases
33	Past year but not past month, or past 3 years but not past year	Missing (lifetime use known)		Missing	0	4,6,17	1. R1 2. R2 3. R3 4. RS1*DS 5. RS1*(1-DS)*PMS	No Cases	No Cases	No Cases
34	Missing (lifetime use known)	Past year but not past month, or past 3 years but not past year	Missing		0	3,5,18	1. R1 2. R2 3. R3 4. RC1*DC 5. RC1*(1-DC)*PMC	No Cases	No Cases	No Cases
35	Past year but not past month, or past 3 years but not past year	Not past year			0	14,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
36	Not past year	Past year but not past month, or past 3 years but not past year			2	13,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	1-3: 0 1,2,4: 0 1,4: 1	1-3: 1	No Cases
37	Past 3 years	Past 3 years	Missing	Missing	0	1,2,5,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/(RC1+RC2+ RC3) 5. RS1*DS/ (RS1+RS2+RS3) 6. RS1*(1-DS)* PMS/(RS1+RS2+ RS3)	No Cases	No Cases	No Cases
38	Not past month	Past year but not past month, or past 3 years but not past year			1	15,18	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	1-3: 0 1,2,4: 0 1,4: 1	No Cases
39	Missing (lifetime use known)	Past 3 years	Missing	Missing	0	2,3,5,6	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2+RC3) 4. RC1*(1-DC)* PMC/(RC1+RC2+ RC3) 5. DS 6. PMS	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
40	Past 3 years	(Past month)	Missing	Missing	0	1,5,6,8	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/(RC1+RC2+RC3) 4. RC1*(1-DC)*PMC/(RC1+RC2+RC3) 5. DS 6. PMS	No Cases	No Cases	No Cases
41	(Past month)	Not past year	Missing		0	5,7,14	1. R3/(R3+R4) 2. DC 3. PMC	No Cases	No Cases	No Cases
42	Past 3 years	Past year but not past month, or past 3 years but not past year	Missing		0	1,5,18	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3.RC1*DC/(RC1+RC2+RC3) 4. RC1*(1-DC)*PMC/(RC1+RC2+RC3)	No Cases	No Cases	No Cases
43	Past year but not past month, or past 3 years but not past year	Past 3 years		Missing	0	2,6,17	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3.RS1*DS/(RS1+RS2+RS3) 4. RS1*(1-DS)*PMS/(RS1+RS2+RS3)	No Cases	No Cases	No Cases
44	Past year but not past month, or past 3 years but not past year	Not past month			0	16,17	1. R2/(R2+R3+R4) 2. R3/(R2+R3+R4)	No Cases	No Cases	No Cases

Table E.34 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Chew Recency	Snuff Recency	Chew 30-Day Freq.	Snuff 30-Day Freq.				12-17	18-25	26+
45	Past Year	Past year but not past month, or past 3 years but not past year	Missing		1	5,11,18	1. R1/(R1+R2+R3) 2. R2/(R1+R2+R3) 3. RC1*DC/ (RC1+RC2) 4.RC1*(1-DC)*PMC/ (RC1+RC2)	1-3: 0 1,2,4: 0 1,4: 1	No Cases	No Cases

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month smokeless tobacco use | lifetime smokeless tobacco use)
2. R2 = P(Past year but not past month smokeless tobacco use | lifetime smokeless tobacco use)
3. R3 = P(Past 3 years but not past year smokeless tobacco use | lifetime smokeless tobacco use)
4. R4 = P(Lifetime but not past 3 years smokeless tobacco use | lifetime smokeless tobacco use)
5. RC1 = P(Past month chewing tobacco use | lifetime chewing tobacco use)
6. RC2 = P(Past year but not past month chewing tobacco use | lifetime chewing tobacco use)
7. RC3 = P(Past 3 years but not past year chewing tobacco use | lifetime chewing tobacco use)
8. RS1 = P(Past month snuff use | lifetime snuff use)
9. RS2 = P(Past year but not past month snuff use | lifetime snuff use)
10. RS3 = P(Past 3 years but not past year snuff use | lifetime snuff use)
11. DC = P(Daily chewing tobacco use | past month chewing tobacco use)
12. DS = P(Daily snuff use | past month snuff use)
13. PMC = P(Chewing tobacco use on a given day in the past month | past month use of chewing tobacco)
14. PMS = P(Snuff use on a given day in the past month | past month use of snuff)

Table E.35 Likeness Constraints for Pipe Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means

Table E.36 Constraints and Portion of the Predictive Mean Vector for Pipe Recency and Frequency

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing (lifetime use known)	3	None	1. R1	No Cases	1,2: 1	1,2: 2

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month pipe use | lifetime pipe use)

Table E.37 Logical Constraints for Various Drugs Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be less than or equal to recipient's maximum possible past year drug frequency of use Recipient's maximum possible drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used drug, as determined by the month of first drug use (2) if the maximum period recipient could have used drug is greater than 30, but recipient is a past month drug user with a nonmissing 30-day drug frequency, the past year drug frequency must be less than or equal to the maximum period (minus the number of days recipient did not use drug in the past month) (3) if recipient is not a past month drug user, the past year drug frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to recipient's minimum possible past year drug frequency of use Recipient's minimum possible drug frequency of use in the past year is limited by the following factors: (1) if recipient is a past month drug user, it must be at least as much as the 30-day drug frequency (2) if recipient is not a past month drug user but is a past year drug user, it must be at least 1

Table E.37 Logical Constraints for Various Drugs Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC3	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to recipient's 30-day drug use
LogC4	Donor's 30-day drug use must be less than or equal to recipient's maximum number of days could have used drug in past 30 days
LogC5	Donor's 30-day drug use must be greater than or equal to recipient's minimum number of days could have used drug in past 30 days
LogC6	Donor's 30-day drug use must be greater than or equal to recipient's DR5DAY (# days had 5+ drinks in past 30 days)
LogC7	Donor's 30-day drug use must be greater than or equal to donor's proportion of past year drug use * recipient's maximum number of days could have used drug in past year [minus 335]
LogC8	Donor must be a past month drug user (drug recency = 1)
LogC9	If recipient's age at first drug use equals his or her current age, donor's 30-day drug frequency (1) cannot be greater than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC10	If recipient's age at first drug use equals his or her current age, (1) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC11	Recipient's estimated 30-day drug frequency is not given/legitimately skipped (estimated drug frequency ≠ 1-6)
LogC12	If recipient's age at first drug use equals his or her current age, (1) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between his or her interview date and date of first drug use (minus 30) and (2) donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year cannot be greater than recipient's days between the interview date and birthday (minus 30)
LogC13	Donor must be a past year but not past month drug user (drug recency = 2)
LogC14	Donor's DR5DAY value must be less than or equal to recipient's 30-day drug frequency
LogC15	If recipient's age at first drug use equals his or her current age, (1) donor's DR5DAY must be less than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) donor's DR5DAY must be less than recipient's days between his or her interview date and birthday (inclusive)
LogC16	Donor must be a past month or past year but not past month drug user (drug recency = 1 or 2)
LogC17	Donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than or equal to donor's 30-day drug frequency
LogC18	If recipient's month/year of first drug use data indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than recipient's 30-day drug frequency

Table E.37 Logical Constraints for Various Drugs Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC19	If recipient's month/year of first drug use data indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than recipient's DR5DAY value
LogC20	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month drug user, then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than donor's 30-day drug frequency
LogC21	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than 1
LogC22	If recipient's month/year of first drug use data indicate that he or she must have used drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month drug user, then donor's proportion of past year drug use * recipient's maximum number of days could have used drug in the past year must be greater than 1

Table E.38 Likeness Constraints for Various Drugs Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.39 Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Mo. Freq.	30-Day Freq.	30-Day Binge Drink				12-17	18-25	26+
1	(Past month)	Missing	Missing		11	1,2,4-11,17,19	1. PM 2. PY	1-3: 0 1,3: 4	1-3: 3	1-3: 3 1,3: 1
2	(Past month)		Missing		195	4-6,8,9,11	1. PM	1-3: 27 1,3: 1	1-3: 100 1,3: 1	1-3: 66
3	(Past month)	Missing			183	1-3,8,10,18	1. PY	1-3: 51	1-3: 68	1-3: 64
4	(Past year but not past month)	Missing			123	1,2,12,13	1. PY	1-3: 67	1-3: 34	1-3: 22
5	(Past month)			Missing	528	8,14,15	1. PMB	1-3: 87	1-3: 255	1-3: 185 1,3: 1
6	(Past month)		Missing	Missing	17	4,5,8,9,11	1. PM 2. PMB	1-3: 7 1,3: 2	1-3: 5 1,3: 1	1-3: 2
7	(Past month)	Missing		Missing	67	1-3,8,10,14, 15,18	1. PY 2. PMB	1-3: 17 1,3: 3	1-3: 18 1,3: 1	1-3: 27 1,3: 1
8	(Past month)	Missing	Missing	Missing	16	1,2,4,5,7-11, 17,20	1. PM 2. PY 3. PMB	1-3: 2 1,3: 4	1-3: 1 1,3: 1	1-3: 4 1,3: 4
9	Past year		Missing	Missing	345	4,5,9,11,16	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. R1*PMB/(R1+R2)	1-3: 55 1,3: 34	1-3: 73 1,3: 51	1-3: 61 1,3: 71
10	Past year	Missing	Missing	Missing	72	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY 4. R1*PMB/(R1+R2)	1-3: 4 1,3: 19	1-3: 2 1,3: 21	1-3: 3 1,3: 23

Table E.39 Constraints and Portion of the Predictive Mean Vector for Alcohol Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Mo. Freq.	30-Day Freq.	30-Day Binge Drink				12-17	18-25	26+
11	Lifetime (known)	Missing	Missing	Missing	436	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY 5. R1*PMB	1-3: 87 1,3: 224	1-3: 7 1,3: 99	1-3: 0 1,3: 19

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month alcohol use | lifetime alcohol use)
2. R2 = P(Past year but not past month alcohol use | lifetime alcohol use)
3. PM = P(Alcohol use on a given day in the past month | past month alcohol use)
4. PY = P(Alcohol use on a given day in the past year | past year alcohol use)
5. PMB = P(Past month binge alcohol use | past month alcohol use)

Table E.40 Constraints and Portion of the Predictive Mean Vector for Inhalants Recency and Frequency

#	Missingness Pattern			Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.	30-Day Freq.				12-17	18-25	26+
1	(Past month)	Missing	Missing	3	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 0 1,3: 3	No Cases	No Cases
2	(Past month)		Missing	6	4,5,8,9,11	1. PM	1-3: 3 1,3: 2	1-3: 1	No Cases
3	(Past month)	Missing		10	1-3,8,10,18	1. PY	1-3: 8 1,3: 1	1-3: 0 1,3: 1	No Cases
4	(Past year but not past month)	Missing		20	1,2,12,13	1. PY	1-3: 19 1,3: 1	No Cases	No Cases
5	Past year		Missing	18	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/(R1+R2)	1-3: 3 1,3: 11	1-3: 0 1,3: 1	1-3: 0 1,3: 3
6	Past year	Missing	Missing	5	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/(R1+R2) 3. PY	1-3: 0 1,3: 3	1-3: 0 1,3: 2	No Cases
7	Missing (lifetime use known)	Missing	Missing	204	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 16 1,3: 150	1-3: 0 1,3: 29	1-3: 0 1,3: 9

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month inhalants use | lifetime inhalants use)
2. R2 = P(Past year but not past month inhalants use | lifetime inhalants use)
3. PM = P(Inhalants use on a given day in the past month | past month inhalants use)
4. PY = P(Inhalants use on a given day in the past year | past year inhalants use)

Table E.41 Constraints and Portion of the Predictive Mean Vector for Marijuana Recency and Frequency

#	Missingness Pattern			Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.	30-Day Freq.				12-17	18-25	26+
1	(Past month)	Missing	Missing	15	1,2,4,5,8-11, 17,20	1. PM 2. PY	1-3: 6 1,3: 3	1-3: 4	1-3: 0 1,3: 2
2	(Past month)		Missing	7	4,5,8,9,11	1. PM	1-3: 5	1-3: 1	1-3: 1
3	(Past month)	Missing		94	1,2,3,8,10,18	1. PY	1-3: 50 1,3: 1	1-3: 34	1-3: 9
4	(Past year but not past month)	Missing		56	1,2,12,13	1. PY	1-3: 30	1-3: 19	1-3: 7
5	Past year		Missing	90	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-3: 21 1,3: 9	1-3: 36 1,3: 6	1-3: 13 1,3: 5
6	Past year	Missing	Missing	132	1,2,4,5,7,9-12,16,17,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-3: 3 1,3: 26	1-3: 16 1,3: 66	1-3: 4 1,3: 17
7	Missing (lifetime use known)	Missing	Missing	230	1,2,4,5,7,9-12, 17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 7 1,3: 121	1-3: 1 1,3: 76	1-3: 0 1,3: 25

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month marijuana use | lifetime marijuana use)
2. R2 = P(Past year but not past month marijuana use | lifetime marijuana use)
3. PM = P(Marijuana use on a given day in the past month | past month marijuana use)
4. PY = P(Marijuana use on a given day in the past year | past year marijuana use)

Table E.42 Constraints and Portion of the Predictive Mean Vector for Heroin Recency and Frequency

#	Missingness Pattern			Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.	30-Day Freq.				12-17	18-25	26+
1	(Past month)	Missing	Missing	0	1,2,4,5,8-11, 17,20	1. PM 2. PY	No Cases	No Cases	No Cases
2	(Past month)		Missing	1	4,5,8,9,11	1. PM	No Cases	1-3: 0 1,3: 1	No Cases
3	(Past month)	Missing		1	1-3,8,10,18	1. PY	No Cases	No Cases	1-3: 0 1,3: 1
4	(Past year but not past month)	Missing		1	1,2,12,13	1. PY	1-3: 0 1,3: 1	No Cases	No Cases
5	Past year		Missing	4	4,5,9,11,16,17	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-3: 2	1-3: 0 1,3: 2	No Cases
6	Past year	Missing	Missing	11	1,2,4,5,7,9-12, 16,17,20	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-3: 0 1,3: 1	1-3: 0 1,3: 7	1-3: 1 1,3: 2
7	Missing (lifetime use known)	Missing	Missing	16	1,2,4,5,7,9-12,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-3: 0 1,3: 5	1-3: 0 1,3: 8	1-3: 0 1,3: 3

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month heroin use | lifetime heroin use)
2. R2 = P(Past year but not past month heroin use | lifetime heroin use)
3. PM = P(Heroin use on a given day in the past month | past month heroin use)
4. PY = P(Heroin use on a given day in the past year | past year heroin use)

Table E.43 Constraints and Portion of the Predictive Mean Vector for Tranquilizers Recency and Frequency

#	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.				12-17	18-25	26+
1	(Past month)	Missing	7	1,2,10,21	1. PY	1-3: 1 1,3: 1	1-3: 2 1,3: 1	1-3: 2
2	(Past year but not past month)	Missing	13	1,2,10	1. PY	1-3: 6 1,3: 1	1-3: 3	1-3: 3
3	Past year		4	16	1. R1/(R1+R2)	1-3: 4	No Cases	No Cases
4	Past year	Missing	8	1,2,10,12,22	1. R1/(R1+R2) 2. PY	No Cases	1-3: 2 1,3: 3	1-3: 0 1,3: 3
5	Missing (lifetime use known)	Missing	116	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 0 1,3: 40	1-3: 8 1,3: 49	1-3: 0 1,3: 19

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month tranquilizers use | lifetime tranquilizers use)
2. R2 = P(Past year but not past month tranquilizers use | lifetime tranquilizers use)
3. PY = P(Tranquilizers use on a given day in the past year | past year tranquilizers use)

Table E.44 Constraints and Portion of the Predictive Mean Vector for Sedatives Recency and Frequency

#	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Recency	12-Month Freq.				12-17	18-25	26+
1	(Past month)	Missing	3	1,2,10,21	1. PY	1-3: 1 1,3: 2	No Cases	No Cases
2	(Past year but not past month)	Missing	4	1,2,10	1. PY	1-3: 0 1,3: 2	1-3: 0 1,3: 2	No Cases
3	Past year		0	16	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	Past year	Missing	2	1,2,10,12,22	1. R1/(R1+R2) 2. PY	No Cases	1-3: 0 1,3: 2	No Cases
5	Missing (lifetime use known)	Missing	31	1,2,10,12	1. R1 2. R2 3. (R1+R2)*PY	1-3: 1 1,3: 19	1-3: 0 1,3: 7	1-3: 0 1,3: 4

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month sedatives use | lifetime sedatives use)
2. R2 = P(Past year but not past month sedatives use | lifetime sedatives use)
3. PY = P(Sedatives use on a given day in the past year | past year sedatives use)

Table E.45 Logical Constraints for Cocaine Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor must be a past month cocaine user (cocaine recency = 1)
LogC2	Donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be less than or equal to recipient's maximum possible past year cocaine frequency of use Recipient's maximum possible cocaine frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used cocaine, as determined by the month of first cocaine use (2) if the maximum period recipient could have used cocaine is greater than 30, but recipient is a past month cocaine user with a nonmissing 30-day cocaine frequency, the past year cocaine frequency must be less than or equal to the maximum period (the number of days recipient did not use cocaine in the past month) (3) if recipient is not a past month cocaine user, the past year cocaine frequency must be less than or equal to the maximum period (minus 30)
LogC3	Donor's proportion of past year cocaine use * recipient's minimum number of days could have used cocaine in the past year must be greater than or equal to recipient's minimum possible past year cocaine frequency of use Recipient's minimum possible cocaine frequency of use in the past year is limited by the following factors: (1) if recipient is a past month cocaine user, it must be at least as much as the 30-day cocaine frequency (2) if recipient is not a past month cocaine user but is a past year cocaine user, it must be at least 1
LogC4	Recipient's proportion of past year cocaine use * maximum number of days could have used cocaine in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC5	Donor's proportion of past year cocaine use * recipient's number of days could have used cocaine in the past year must be greater than or equal to 30-day cocaine use
LogC6	Donor's 30-day cocaine use must be less than number of days between recipient's interview date and birthday (inclusive)
LogC7	Donor's 30-day cocaine use must be less than recipient's maximum number of days could have used cocaine in past 30 days
LogC8	Donor's 30-day cocaine use must be greater than recipient's minimum number of days could have used cocaine in past 30 days
LogC9	If recipient's age at first cocaine use equals his or her current age, donor's 30-day cocaine frequency (1) cannot be greater than recipient's days between his or her interview date and date of first cocaine use (inclusive) and (2) cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC10	If recipient's age at first cocaine use equals his or her current age, (1) donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year cannot be greater than recipient's days between his or her interview date and date of first drug use (inclusive) and (2) donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC11	Recipient's estimated cocaine 30-day frequency is not given/legitimately skipped (estimated cocaine frequency ≠ 1-6)
LogC12	Donor's crack recency equals recipient's crack recency
LogC13	Donor must be a past year but not past month cocaine user (cocaine recency = 2)
LogC14	If recipient's age at first cocaine use equals his or her current age, donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year cannot be greater than recipient's days between his or her interview date and date of first cocaine use (minus 29)
LogC15	Donor must be a past month or past year but not past month cocaine user (cocaine recency = 1 or 2)
LogC16	Donor must be a past month, past year but not past month, or a lifetime but not past year cocaine user (cocaine recency = 1, 2, or 3)
LogC17	If recipient's age at first cocaine use equals his or her current age, donor cannot be a lifetime but not past year cocaine user (cocaine recency ≠ 3)
LogC18	Donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be less than or equal to recipient's maximum possible past year crack frequency of use Recipient's maximum possible crack frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used crack, as determined by the month of first crack use (2) if the maximum period recipient could have used crack is greater than 30, but recipient is a past month crack user with a nonmissing 30-day crack frequency, the past year crack frequency must be less than or equal to the maximum period (the number of days recipient did not use in the past month) (3) if recipient is not a past month crack user, the past year crack frequency must be less than or equal to the maximum period (minus 30)
LogC19	Donor's proportion of past year crack use * recipient's minimum number of days could have used crack in the past year must be greater than or equal to recipient's minimum possible past year crack frequency of use Recipient's minimum possible crack frequency of use in the past year is limited by the following factors: (1) if recipient is a past month crack user, it must be at least as much as the 30-day crack frequency (2) if recipient is not a past month crack user but is a past year crack user, it must be at least 1
LogC20	Recipient's proportion of past year crack use * maximum number of days could have used crack in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC21	Donor's proportion of past year crack use * recipient's number of days could have used crack in the past year must be greater than or equal to 30-day crack use
LogC22	Donor's 30-day crack use must be less than number of days between recipient's interview date and birthday (inclusive)
LogC23	Donor's 30-day crack use must be less than recipient's maximum number of days could have used crack in past 30 days

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC24	Donor's 30-day crack use must be greater than recipient's minimum number of days could have used crack in past 30 days
LogC25	If recipient's age at first crack use equals his or her current age, donor's 30-day crack frequency (1) cannot be greater than recipient's days between his or her interview date and date of first crack use (inclusive) and (2) cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC26	If recipient's age at first crack use equals his or her current age, (1) donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and date of first crack use (inclusive) and (2) donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC27	Recipient's estimated 30-day crack frequency is not given/legitimately skipped (estimated crack frequency ≠ 1-6)
LogC28	Donor must be a past month crack user (crack recency = 1)
LogC29	Donor must be a past month or past year not past month crack user (crack recency = 1, 2)
LogC30	Donor must be a past month, past year not past month, or lifetime but not past year crack user (crack recency = 1, 2)
LogC31	Donor's cocaine recency must equal recipient's cocaine recency, or donor's cocaine recency must equal recipient's cocaine recency (minus 10)
LogC32	If recipient's age at first crack use equals his or her current age, donor cannot be a lifetime but not past year crack user (crack recency ≠ 3)
LogC33	Donor must be a past year but not past month crack user (crack recency = 2)
LogC34	If recipient's age at first crack use equals his or her current age, donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year cannot be greater than recipient's days between his or her interview date and date of first crack use (minus 29)
LogC35	If recipient's month/year of first use data for cocaine indicate that he or she must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than recipient's 30-day cocaine frequency
LogC36	If recipient's month/year of first use data for cocaine indicate that he or she must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than donor's 30-day cocaine frequency
LogC37	If recipient's month/year of first use data for cocaine indicate that he or she must have used cocaine at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month cocaine user, then donor's proportion of past year cocaine use * recipient's maximum number of days could have used cocaine in the past year must be greater than donor's 30-day cocaine frequency
LogC38	If recipient's month/year of first use data for crack indicate that he or she must have used crack at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be greater than recipient's 30-day crack frequency

Table E.45 Logical Constraints for Cocaine Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC39	If recipient's month/year of first use data for crack indicate that he or she must have used crack at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year crack use * recipient's maximum number of days could have used crack in the past year must be greater than donor's 30-day crack frequency
LogC40	If recipient's month/year of first use data for crack indicate that he or she must have used crack at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month crack user, then donor's proportion of past year crack use * recipient's maximum number of days could have used in the past year must be greater than donor's 30-day crack frequency

Table E.46 Likeness Constraints for Cocaine Recency and Frequency

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]
LikC4	Donor's crack recency agrees with recipient's crack recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
1	(Past month)		Missing		Missing		10	1-12,36	1. PM 2. PY	No Cases	1-4: 0 1-3: 0 1,3: 3	1-4: 1 1-3: 0 1,3: 4 3: 2
2	(Past month)				Missing		10	1,6-9,11,12	1. PM	1-4: 0 1-3: 1 1,3: 1	1-4: 1 1-3: 1 1,3: 3 3: 1	1-4: 0 1-3: 1 1,3: 0 3: 1
3	(Past month)		Missing				5	2-4,10,12,35	1. PY	No Cases	1-4: 1	1-4: 1 1-3: 1 1,3: 1 3: 0 None: 0 None*: 1
4	(Past year not past month)		Missing				17	2-4,12-14	1. PY	No Cases	1-4: 7 1-3: 3 1,3: 1	1-4: 2 1-3: 2 1,3: 2
5	Past year				Missing		8	6-9,11,12,15	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-4: 6 1-3: 0 1,3: 1	1-4: 1
6	Past year		Missing		Missing		5	2-12,15,37	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 4
7	Missing (lifetime use known)		Missing		Missing		62	2-12,16,17	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 16	1-4: 1 1-3: 0 1,3: 41	1-4: 0 1-3: 0 1,3: 4
8	(Past month)	(Past month)		Missing		Missing	0	1,18-27,39	1. PM 2. PY	No Cases	No Cases	No Cases

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
9	(Past month)	(Past month)				Missing	0	1,22-25,27,28	1. PM	No Cases	No Cases	No Cases
10	(Past month)	(Past month)		Missing			0	15,18-20,26, 28,38	1. PY	No Cases	No Cases	No Cases
11	(Past year not missing)	(Past year not past month)		Missing			1	15,18-20,26,29	1. PY	No Cases	1-4: 0 1-3: 0 1,3: 1	No Cases
12	(Past month)	Past year				Missing	0	1,22-25,27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
13	(Past month)	Past year		Missing		Missing	0	1,18-27,29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
14	(Past month)	Missing (lifetime use known)		Missing		Missing	2	16,18-26,30-32	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	1-4: 0 1-3: 0 1,3: 2	No Cases
15	(Past month)	(Past month)	Missing	Missing			0	1-4,10,18-20, 26,28,35,38	1. PY	No Cases	No Cases	No Cases
16	(Past month)	(Past year but not past month)	Missing	Missing			0	1-4,10,18-20, 26,33,35	1. PY	No Cases	No Cases	No Cases
17	(Past year but not past month)	(Past year but not past month)	Missing	Missing			1	2-4,14,18-20, 33,34	1. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
18	(Past month)	(Past month)			Missing	Missing	0	1,6-9,11,22-25, 27,28	1. PM	No Cases	No Cases	No Cases
19	(Past month)	(Past month)	Missing	Missing	Missing	Missing	0	1-11,18-28,36, 39	1. PM 2. PY	No Cases	No Cases	No Cases

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
20	(Past month)	(Past month)	Missing		Missing	Missing	0	1-11,16,22-25, 27,28,36	1. PM 2. PY	No Cases	No Cases	No Cases
21	(Past month)	(Past month)		Missing	Missing	Missing	0	1,6-9,11,18-28, 39	1. PM 2. PY	No Cases	No Cases	No Cases
22	(Past month)	(Past month)	Missing	Missing	Missing		0	1-11,18-21,26, 28,36,38	1. PM 2. PY	No Cases	No Cases	No Cases
23	(Past month)	(Past year not past month)	Missing	Missing	Missing		0	1-11,18-20,33, 34,36	1. PM 2. PY	No Cases	No Cases	No Cases
24	(Past month)	(Past month)	Missing	Missing		Missing	0	1-4,10,18-26, 28,36	1. PM 2. PY	No Cases	No Cases	No Cases
25	(Past month)	(Past month)		Missing	Missing		0	1,6-9,18-20,26, 28,38	1. PM 2. PY	No Cases	No Cases	No Cases
26	(Past month)	(Past year not past month)		Missing	Missing		0	1,6-9,11,18-20, 26,33	1. PM 2. PY	No Cases	No Cases	No Cases
27	(Past month)	(Past month)	Missing			Missing	0	1-4,10,22-25, 27,28,35	1. PM 2. PY	No Cases	No Cases	No Cases
28	Past year	Past year			Missing	Missing	1	6-9,11,15, 22-25,27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
29	Past year	Past year	Missing		Missing	Missing	0	2-11,15,21-25, 27,29,37	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
30	Past year	Past year		Missing	Missing	Missing	1	6-9,11,15, 18-27,29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	1-4: 0 1-3: 0 1,3: 1
31	Past year	Past year	Missing	Missing	Missing	Missing	4	2-11,15,18-27, 29,37,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-4: 0 1-3: 0 1,3: 3	1-4: 0 1-3: 0 1,3: 1

Table E.47 Constraints and Portion of the Predictive Mean Vector for Cocaine Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Cocaine Recency	Crack Recency	Cocaine 12-Mo. Freq.	Crack 12-Mo. Freq.	Cocaine 30-Day Freq.	Crack 30-Day Freq.				12-17	18-25	26+
32	Past year	Missing (lifetime use known)		Missing	Missing	Missing	1	1,6-9,11,15, 18-27,30	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-4: 0 1-3: 0 1,3: 1	No Cases	No Cases
33	Past year	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	2-11,15,18-27, 30,32,37	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
34	(Past month)	Missing (lifetime use known)		Missing	Missing	Missing	1	1,6-9,11,18-27, 30,32	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	1-4: 0 1-3: 0 1,3: 1	No Cases
35	(Past month)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	0	1-11,18-27,30, 36	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
36	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	Missing	Missing	7	2-11,16-27,30	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-4: 0 1-3: 0 1,3: 1	1-4: 0 1-3: 0 1,3: 3	1-4: 0 1-3: 0 1,3: 3
37	(Past month)	Past year		Missing	Missing	Missing	0	1,6-9,11,18-27, 29,40	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
38	(Past month)	Past year			Missing	Missing	0	1,6-9,22-25, 27,29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases

*A donor could not be found for these cases. Imputed values were randomly assigned within the bounds created during the editing process.

NOTE: Cocaine users included crack users and cocaine users who were not crack users.

[1] The predictive mean vector components are defined by the following:

1. R1 = P(Past month cocaine use | lifetime cocaine use)
2. R2 = P(Past year but not past month cocaine use | lifetime cocaine use)
3. PM = P(Cocaine use on a given day in the past month | past month use of cocaine)
4. PY = P(Cocaine use on a given day in the past year | past year use of cocaine)

Table E.48 Logical Constraints for Hallucinogens Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year hallucinogen use * recipient's maximum number of days could have used hallucinogens in the past year must be less than or equal to recipient's maximum possible past year hallucinogen frequency of use Recipient's maximum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used hallucinogens, as determined by the month of first hallucinogen use (2) if the maximum period recipient could have used hallucinogens is greater than 30, but recipient is a past month hallucinogen user with a nonmissing 30-day hallucinogen frequency, the past year hallucinogen frequency must be less than or equal to the maximum period (the number of days recipient did not use hallucinogens in the past month) (3) if recipient is not a past month hallucinogen user, the past year hallucinogen frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year hallucinogen use * recipient's minimum number of days could have used hallucinogens in the past year must be greater than or equal to recipient's minimum possible past year hallucinogen frequency of use Recipient's minimum possible hallucinogen frequency of use in the past year is limited by the following factors: (1) if recipient is a past month hallucinogen user, it must be at least as much as the 30-day hallucinogen frequency (2) if recipient is not a past month hallucinogen user but is a past year hallucinogen user, it must be at least 1
LogC3	Recipient's proportion of past year hallucinogen use * maximum number of days could have used hallucinogens in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC4	Donor's 30-day hallucinogen use must be less than the number of days between recipient's interview date and birthday (inclusive)
LogC5	Donor's 30-day hallucinogen use must be less than recipient's maximum number of days could have used hallucinogens in past 30 days
LogC6	Donor's 30-day hallucinogen use must be greater than recipient's minimum number of days could have used hallucinogens in past 30 days
LogC7	Donor must be an LSD user (LSD recency $\neq$ 91)
LogC8	Donor must be a PCP user (PCP recency $\neq$ 91)
LogC9	Donor must be an Ecstasy user (Ecstasy recency $\neq$ 91)
LogC10	Donor's LSD recency must equal recipient's LSD recency
LogC11	Donor's PCP recency must equal recipient's PCP recency
LogC12	Donor's Ecstasy recency must equal recipient's Ecstasy recency
LogC13	Donor must be an LSD and PCP user (LSD and PCP recencies $\neq$ 91)
LogC14	Donor must be an LSD and Ecstasy user (LSD and Ecstasy recencies $\neq$ 91)
LogC15	Donor must be a PCP and Ecstasy user (PCP and Ecstasy recencies $\neq$ 91)
LogC16	Donor must be an LSD and PCP and Ecstasy user (LSD and PCP and Ecstasy recencies $\neq$ 91)

Table E.48 Logical Constraints for Hallucinogens Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC17	Donor must be a past month hallucinogens user (hallucinogen recency = 1)
LogC18	Donor must be a hallucinogen past year but not past month or past month hallucinogen user (hallucinogen recency = 1 or 2)
LogC19	Donor must be a hallucinogen user (hallucinogen recency = 1, 2, or 3)
LogC20	Donor must be an LSD past year but not past month or past month LSD user (LSD recency = 1 or 2)
LogC21	Donor must be a PCP past year but not past month or past month PCP user (PCP recency = 1 or 2)
LogC22	Donor must be an Ecstasy past year but not past month or past month Ecstasy user (Ecstasy recency = 1 or 2)
LogC23	Donor must not be an LSD past year but not past month or past month LSD user (LSD recency ≠ 1 or 2)
LogC24	Donor must not be a PCP past year but not past month or past month PCP user (PCP recency ≠ 1 or 2)
LogC25	Donor must not be an Ecstasy past year but not past month or past month Ecstasy user (Ecstasy recency ≠ 1 or 2)
LogC26	Donor's hallucinogen recency must equal recipient's hallucinogen recency, or donor's hallucinogen recency must equal recipient's hallucinogen recency (minus 10)
LogC27	If recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year hallucinogens use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than recipient's 30-day hallucinogen frequency
LogC28	If recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year hallucinogens use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than donor's 30-day hallucinogen frequency
LogC29	If recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month hallucinogens user, then donor's proportion of past year use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than donor's 30-day hallucinogen frequency
LogC30	If recipient is a past month hallucinogens user and recipient's month/year of first use data for hallucinogens indicate that he or she must have used hallucinogens at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year hallucinogens use * recipient's maximum number of days could have used hallucinogens in the past year must be greater than donor's 30-day hallucinogen frequency
LogC31	If recipient has never used hallucinogens other than LSD, PCP, and Ecstasy, then donor must not have recency values that would cause recipient to have imputation-revised recency for overall hallucinogens less than the minimum of the imputation-revised recencies for LSD, PCP, and Ecstasy

Table E.49 Likeness Constraints for Hallucinogens Recency and Frequency

Constraint #	Likeness Constraint
LikC1	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]
LikC2	Donor's indicator of lifetime use of other hallucinogens = Recipient's indicator of lifetime use of other hallucinogens
LikC3	State rank of donor = State rank of recipient
LikC4	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC5	Donor's recencies for LSD, PCP, and Ecstasy agree with recipient's recencies for LSD, PCP, and Ecstasy (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
1		Missing (lifetime use known)					3	7,11,12,26	1. R1 2. R2	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 0 1-3: 1	No Cases
2			Missing (lifetime use known)				1	8,10,12,26	1. R1 2. R2	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
3		Missing (lifetime use known)	Missing (lifetime use known)				1	7,8,12,26,31	1. R1 2. R2	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
4	(Past month)				Missing	Missing	5	1-6,17,28	1. PM 2. PY	1-5: 0 1-4: 0 1-3: 3	1-5: 1 1-4: 0 1-3: 1	No Cases
5	(Past month)					Missing	8	4-6,17	1. PM	1-5: 2 1-4: 0 1-3: 3	1-5: 2	1-5: 1
6	(Past year)				Missing		21	1-3,18,30	1. PY	1-5: 3 1-4: 2 1-3: 3 1,2: 1	1-5: 6 1-4: 4 1-3: 1	1-5: 0 1-4: 0 1-3: 1
7	(Past month)	Missing (lifetime use known)				Missing	0	4-7,11,12,17	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
8	(Past month)		Missing (lifetime use known)			Missing	1	4-6,8,10,12, 17	1. R1 2. R2 3. PM	1-5: 0 1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases
9	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)			Missing	0	4-8,12,17,31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
10	(Past year)	Missing (lifetime use known)			Missing		0	1-3,7,11,12, 18,30	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
11	(Past year)		Missing (lifetime use known)		Missing		1	1-3,8,10,12, 18,30	1. R1 2. R2 3. PY	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
12	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)		Missing		0	1-3,7,8,12,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
13	(Past month)	Missing (lifetime use known)			Missing	Missing	0	1-6,7,11,12, 17,28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
14	(Past month)		Missing (lifetime use known)		Missing	Missing	0	1-6,8,10,12, 17,28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
15	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	Missing	0	1-8,12,17,28, 31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
16	Past year	(Not past month)	(Not past month)	(Not past month)		Missing	12	4-6,10-12,18	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 1 1-4: 0 1-3: 1	1-5: 0 1-4: 5 1-3: 2	1-5: 0 1-4: 0 1-3: 3
17	Past year	(Not past month)	(Not past month)	(Not past month)	Missing	Missing	6	1-6,10-12,18, 29	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 5	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
18	Past year	Past year	(Not past month)	(Not past month)		Missing	10	4-6,11,12,18, 20,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 1 1-3: 6	1-5: 0 1-4: 0 1-3: 3	No Cases
19	Past year	(Not past month)	Past year	(Not past month)		Missing	1	4-6,10,12,18, 21,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
20	Past year	Past year	Past year	(Not past month)		Missing	0	4-6,12,18,20, 21,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
21	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)		Missing	4	4-7,11,12,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 3	1-5: 0 1-4: 0 1-3: 1	No Cases
22	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)		Missing	1	4-6,8,10,12, 18,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
23	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)		Missing	0	4-8,12,18,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
24	Past year	Past year	(Not past month)	(Not past month)	Missing	Missing	0	1-6,11,12,18, 20,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
25	Past year	(Not past month)	Past year	(Not past month)	Missing	Missing	0	1-6,10,12,18, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
26	Past year	Past year	Past year	(Not past month)	Missing	Missing	0	1-6,12,18,20, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
27	Past year	Missing (lifetime use known)	(Not past month)	(Not past month)	Missing	Missing	0	1-7,11,12,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
28	Past year	(Not past month)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0	1-6,8,11,12, 18,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
29	Past year	Missing (lifetime use known)	Missing (lifetime use known)	(Not past month)	Missing	Missing	0	1-8,12,18,29, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
30	Missing (lifetime use known)	(Not past year)	(Not past year)	(Not past year)	Missing	Missing	41	1-6,10-12,19	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 24	1-5: 0 1-4: 0 1-3: 14	1-5: 0 1-4: 0 1-3: 3
31	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing	Missing	37	1-7,11,12,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 10	1-5: 0 1-4: 0 1-3: 19	1-5: 0 1-4: 0 1-3: 8
32	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	(Not past year)	Missing	Missing	19	1-6,8,10,12, 19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 4	1-5: 0 1-4: 0 1-3: 7	1-5: 0 1-4: 0 1-3: 8
33	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing	Missing	0	1-8,12,19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	No Cases	No Cases	No Cases
34				Missing (lifetime use known)			2	9-11,26	1. R1 2. R2	No Cases	1-5: 0 1-4: 1 1-3: 1	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
35		Missing (lifetime use known)		Missing (lifetime use known)			2	7,9,11,26,31	1. R1 2. R2	1-5: 0 1-4: 0 1-3: 2	No Cases	No Cases
36			Missing (lifetime use known)	Missing (lifetime use known)			0	8-10,26,31	1. R1 2. R2	No Cases	No Cases	No Cases
37		Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)			0	7-9,26,31	1. R1 2. R2	No Cases	No Cases	No Cases
38	(Past month)			Missing (lifetime use known)		Missing	0	4-6,9-11,17	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
39	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)		Missing	0	4-7,9,11,17, 31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
40	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-6,8-10,17, 31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
41	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-9,17,31	1. R1 2. R2 3. PM	No Cases	No Cases	No Cases
42	(Past year)			Missing (lifetime use known)	Missing		0	1-3,9-11,18, 30	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
43	(Past year)	Missing (lifetime use known)		Missing (lifetime use known)	Missing		0	1-3,7,9,11,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
44	(Past year)		Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	1-3,8-10,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
45	(Past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing		0	1-3,7-9,18, 30,31	1. R1 2. R2 3. PY	No Cases	No Cases	No Cases
46	(Past month)			Missing (lifetime use known)	Missing	Missing	0	1-6,9-11,17, 28	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
47	(Past month)	Missing (lifetime use known)		Missing (lifetime use known)	Missing	Missing	0	1-7,9,11,17, 28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
48	(Past month)		Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-6,8-10,17, 28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
49	(Past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-9,17,28,31	1. R1 2. R2 3. PM 4. PY	No Cases	No Cases	No Cases
50	Past year	(Not past month)	(Not past month)	Past year		Missing	8	4-6,10,11,18, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 5	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 2

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
51	Past year	Past year	(Not past month)	Past year		Missing	3	4-6,11,18,20, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 2	1-5: 0 1-4: 0 1-3: 1	No Cases
52	Past year	(Not past month)	Past year	Past year		Missing	0	4-6,10,18,21, 22,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
53	Past year	Past year	Past year	Past year		Missing	0	4-6,18,20-22, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
54	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)		Missing	6	4-6,9-11,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	1-5: 0 1-4: 0 1-3: 3	1-5: 0 1-4: 2 1-3: 1	No Cases
55	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)		Missing	2	4-7,9,11,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	1-5: 0 1-4: 0 1-3: 2	No Cases
56	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-6,8-10,18, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
57	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)		Missing	0	4-9,18,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
58	Past year	(Not past month)	(Not past month)	Past year	Missing	Missing	2	1-6,10,11,18, 22,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-5: 0 1-4: 0 1-3: 2	No Cases
59	Past year	Past year	(Not past month)	Past year	Missing	Missing	2	1-6,11,18,20, 22,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	1-5: 0 1-4: 0 1-3: 2	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
60	Past year	(Not past month)	Past year	Past year	Missing	Missing	1	1-6,10,18,21, 22,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
61	Past year	Past year	Past year	Past year	Missing	Missing	0	1-6,18,20-22, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
62	Past year	(Not past month)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	1-6,9-11,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
63	Past year	Missing (lifetime use known)	(Not past month)	Missing (lifetime use known)	Missing	Missing	0	1-7,9,11,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
64	Past year	(Not past month)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	1	1-6,8-10,18, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
65	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	0	1-9,18,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
66	Missing (lifetime use known)	(Not past year)	(Not past year)	Missing (lifetime use known)	Missing	Missing	67	1-6,9-11,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 24	1-5: 0 1-4: 0 1-3: 33	1-5: 0 1-4: 0 1-3: 10
67	Missing (lifetime use known)	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing	Missing	12	1-7,9,11,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 3	1-5: 0 1-4: 0 1-3: 8	1-5: 0 1-4: 0 1-3: 1

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
68	Missing (lifetime use known)	(Not past year)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	3	1-6,8-10,19, 31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 2	No Cases	1-5: 0 1-4: 0 1-3: 1
69	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	3	1-9,19,31	1. R1 2. R2 3. R1*PM 4. (R1+R2)*PY	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 1	1-5: 0 1-4: 0 1-3: 1
70				Past year			0	10,11,22,26	1. R1/(R1+R2)	No Cases	No Cases	No Cases
71		Past year	Past year				0	12,20,21,26, 31	1. R1/(R1+R2)	No Cases	No Cases	No Cases
72		Past year		Past year			0	11,20,22,26, 31	1. R1/(R1+R2)	No Cases	No Cases	No Cases
73	(Past month)	Past year			Missing	Missing	0	1-6,11,12,17, 20,28	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
74	(Past month)		Past year		Missing	Missing	0	1-6,10,12,17, 21,28	1. R1/(R1+R2) 2. PM 3. PY	No Cases	No Cases	No Cases
75	Past year	Missing (lifetime use known)	Past year	Missing (lifetime use known)	Missing	Missing	0	1-7,9,18,21, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
76	Past year	(Not past month)	Past year	Missing (lifetime use known)	Missing	Missing	0	1-6,9,10,18, 21,29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
77			Past year				0	10,12,21,26	1. R1/(R1+R2)	No Cases	No Cases	No Cases
78		Past year					4	11,12,20,26	1. R1/(R1+R2)	1-5: 0 1-4: 1 1-3: 2	1-5: 0 1-4: 1	No Cases
79	(Past month)	Past year				Missing	0	4-7,11,12,17	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases

Table E.50 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Recency and Frequency (continued)

#	Missingness Pattern						Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Hallu-cinogens Recency	LSD Recency	PCP Recency	Ecstasy Recency	Hallu-cinogen 12-Mo. Freq.	Hallu-cinogen 30-Day Freq.				12-17	18-25	26+
80	(Past month)		Past year			Missing	0	4-6,8,10,12, 17	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
81	(Past month)			Past year		Missing	0	4-6,9-11,17	1. R1/(R1+R2) 2. PM	No Cases	No Cases	No Cases
82	(Past month)	Past year			Missing		1	1-3,7,11,12, 17,27	1. R1/(R1+R2) 2. PY	1-5: 0 1-4: 0 1-3: 1	No Cases	No Cases
83	(Past month)		Past year		Missing		0	1-3,8,10,12, 17,27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
84	(Past month)			Past year	Missing		1	1-3,9-11,17, 27	1. R1/(R1+R2) 2. PY	No Cases	1-5: 0 1-4: 0 1-3: 1	No Cases
85	Past year	Missing (lifetime use known)	Missing (lifetime use known)	Past year		Missing	0	4-6,13,18,22, 31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2)	No Cases	No Cases	No Cases
86	Past year	Missing (lifetime use known)	Past year		Missing	Missing	0	1-7,12,18,21, 29,31	1. R1/(R1+R2) 2. R1*PM/ (R1+R2) 3. PY	No Cases	No Cases	No Cases
87	Past month	Past year	Past year	Past year	Missing		0	1-3,7-12,16, 17,27	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
88			Past year	Past year			0	11,12,21,22, 26,31	1. R1/(R1+R2)	No Cases	No Cases	No Cases

NOTE: Hallucinogen users included users of LSD, users of PCP, and users of Ecstasy.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month hallucinogens use | lifetime hallucinogens use)
2. R2 = P(Past year but not past month hallucinogens use | lifetime hallucinogens use)
3. PM = P(Hallucinogens use on a given day in the past month | past month hallucinogens use)
4. PY = P(Hallucinogens use on a given day in the past year | past year hallucinogens use)

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency

Constraint #	Logical Constraint
LogC1	Donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be less than or equal to recipient's maximum possible past year parent drug frequency of use Recipient's maximum possible parent drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used parent drug, as determined by the month of first parent drug use (2) if recipient is not a past month parent drug user, the past year parent drug frequency must be less than or equal to the maximum period (minus 30)
LogC2	Donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than or equal to recipient's minimum possible past year parent drug frequency of use (For these drugs, the minimum possible past year parent drug frequency of use is always 1.)
LogC3	Recipient's proportion of past year parent drug use * maximum number of days could have used parent drug in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC4	Donor must be a past month parent drug user (parent drug recency = 1)
LogC5	If recipient's age at first parent drug use equals his or her current age, (1) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and date of first parent drug use (inclusive) and (2) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC6	Donor must be a past year but not past month parent drug user (parent drug recency = 2)
LogC7	If recipient's age at first parent drug use equals his or her current age, (1) donor's proportion of past year parent drug use* recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and date of first parent drug use (minus 30) and (2) donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (minus 30)
LogC8	Donor must be a past month or past year but not past month parent drug user (parent drug recency = 1 or 2)
LogC9	Donor's parent drug recency must equal recipient's parent drug recency, or donor's parent drug recency must equal recipient's parent drug recency (minus 10)
LogC10	Donor must be a past month, past year but not past month, or lifetime but not past year child drug user (child drug recency = 1, 2, or 3)
LogC11	If the number of days between recipient's interview and birthday (inclusive) is between 0 and 30, child drug recency must not equal 2 or 3
LogC12	If the number of days between recipient's interview and birthday (inclusive) is between 0 and 365, child drug recency must not equal 3

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC13	If recipient's age at first parent drug use equals his or her current age, or recipient's age at first child drug use equals his or her current age, or recipient's number of days between his or her interview date and date at first child drug use is less than 30, then donor's parent drug recency must not equal 3
LogC14	Donor must be a past month or past year but not past month child drug user (child drug recency = 1 or 2)
LogC15	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be less than or equal to recipient's maximum possible past year child drug frequency of use Recipient's maximum possible child drug frequency of use in the past year is limited by the following factors: (1) it must be less than or equal to the maximum period recipient could have used child drug, as determined by the month of first child drug use (2) if recipient is not a past month child drug user, the past year child drug frequency must be less than or equal to the maximum period (minus 30)
LogC16	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be greater than or equal to recipient's minimum possible past year child drug frequency of use (For these drugs, the minimum possible past year child drug frequency of use is always 1.)
LogC17	Recipient's proportion of past year child drug use * maximum number of days could have used child drug in the past year must be less than or equal to the number of days between recipient's interview date and birthday (inclusive)
LogC18	If recipient's age at first child drug use equals his or her current age, (1) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and date of first child drug use (inclusive) and (2) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (inclusive)
LogC19	If recipient's age at first child drug use equals his or her current age, (1) donor's proportion of past year child drug use* recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and date of first child drug use (minus 30) and (2) donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than recipient's days between his or her interview date and birthday (minus 30)
LogC20	Donor must be a past month child drug user (child drug recency = 1)
LogC21	Donor must be a past year but not past month child drug user (child drug recency = 2)
LogC22	Donor must be a past month, past year but not past month, or lifetime but not past year parent drug user (parent drug recency = 1, 2, or 3)
LogC23	If recipient's month/year of first use data for the parent drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than 1

Table E.51 Logical Constraints for Stimulants and Pain Relievers Recency and Frequency (continued)

Constraint #	Logical Constraint
LogC24	If recipient's month/year of first use data for the child drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), then donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year must be greater than 1
LogC25	If recipient's month/year of first use data for the parent drug indicate that he or she must have used at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month parent drug user, then donor's proportion of past year parent drug use * recipient's maximum number of days could have used parent drug in the past year must be greater than 1
LogC26	If recipient's month/year of first use data for the child drug indicate that he or she must have used child drug at least once in the interval (1 year before interview, 30 days before interview), and if donor is a past month child drug user, then donor's proportion of past year use * recipient's maximum number of days could have used in the past year must be greater than 1
LogC27	If recipient is not a lifetime user of any type of the parent drug except for the child drug, then donor must not have used parent drug more recently than recipient has used child drug
LogC28	If recipient is not a lifetime user of any type of the parent drug except for the child drug, then donor must not have used parent drug more recently than donor has used child drug
LogC29	Donor's proportion of past year child drug use * recipient's maximum number of days could have used child drug in the past year cannot be greater than donor's proportion of past year parent drug use * recipient's maximum number of days could have used child drug in the past year

Table E.52 Likeness Constraints for Stimulants and Pain Relievers Recency and Frequency

Constraint #	Likeness Constraint
LikC1	Donor's drug recency must match recipient's drug recency (when nonmissing)[1]
LikC2	State rank of donor = State rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor's methamphetamine recency agrees with recipient's methamphetamine recency (when nonmissing)
LikC5	Donor's OxyContin recency agrees with recipient's OxyContin recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
1	(Past month)		Missing		4	1-5,23	1. PY	1-4: 1	1-4: 2	1-4: 0 1-3: 0 1,2: 1
2	(Past year but not past month)		Missing		3	1-3,6,7	1. PY	1-4: 2	No Cases	1-4: 1
3	Past year				0	8,27	1. R1/(R1+R2)	No Cases	No Cases	No Cases
4	Past year		Missing		3	1-3,5,7,8,25,27	1. R1/(R1+R2) 2. PY	No Cases	1-4: 0 1-3: 0 1,2: 2	1-4: 0 1-3: 0 1,2: 1
5	Missing (lifetime use known)		Missing		53	1,2,5,7,22,25,27	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 21	1-4: 1 1-3: 0 1,2: 27	1-4: 0 1-3: 0 1,2: 4
6	(Past month)	(Past month)		Missing	1	4,15-18,20,24	1. PY	No Cases	1-4: 0 1-3: 0 1,2: 1	No Cases
7	(Past year not missing)	(Past year not past month)		Missing	0	15-17,19,21	1. PY	No Cases	No Cases	No Cases
8	(Past month)	Past year			0	14	1. R1/(R1+R2)	No Cases	No Cases	No Cases
9	(Past month)	Past year	Missing		0	1-3,5,7,14,23	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
10	(Past month)	Past year		Missing	2	14-19,26	1. R1/(R1+R2) 2. PY	No Cases	1-4: 0 1-3: 0 1,2: 2	No Cases
11	(Past month)	Past year	Missing	Missing	0	1-3,5,7,14-19, 23,26,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
12	(Past year not missing)	Missing (lifetime use known)		Missing	1	10,15,16,18,19, 22,29	1. R1 2. R2 3. (R1+R2)*PY	No Cases	1-4: 0 1-3: 0 1,2: 1	No Cases

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
13	(Past month)	(Past month)	Missing	Missing	4	1-5,15-18,20, 23,24,29	1. PY	1-4: 0 1-3: 0 1,2: 1	No Cases	1-4: 1 1-3: 0 1,2: 2
14	(Past month)	(Past year not past month)	Missing	Missing	0	1-5,15-17,19, 21,23,29	1. PY	No Cases	No Cases	No Cases
15	(Past year not past month)	(Past year not past month)	Missing	Missing	3	1-3,6,7,15-17, 19,21,29	1. PY	1-4: 1 1-3: 0 1,2: 1	1-4: 0 1-3: 0 1,2: 1	No Cases
16	Past year	Past year			1	8,14,28	1. R1/(R1+R2)	No Cases	1-4: 1	No Cases
17	Past year	Past year	Missing		0	1-35,7,8,14, 25,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
18	Past year	Past year		Missing	2	8,14-19,19, 26,28	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 2	No Cases	No Cases
19	Past year	Past year	Missing	Missing	1	1-3,5,7,8,14-19, 25,26,28,29	1. R1/(R1+R2) 2. PY	No Cases	1-4: 0 1-3: 0 1,2: 1	No Cases
20	Past year	Missing (lifetime use known)		Missing	3	8,10,15,16,18, 19,28,29	1. R1/(R1+R2) 2. PY	1-4: 0 1-3: 0 1,2: 2	1-4: 0 1-3: 0 1,2: 1	No Cases
21	Past year	Missing (lifetime use known)	Missing	Missing	0	1-3,5,7,8,10,12, 15,16,18,19,25,2 8,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
22	(Past month)	Missing (lifetime use known)		Missing	0	4,10,15,16,18, 19,29	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
23	(Past month)	Missing (lifetime use known)	Missing	Missing	1	1-5,10,15,16,18, 19,23,29	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 1	No Cases	No Cases

Table E.53 Constraints and Portion of the Predictive Mean Vector for Stimulants Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency	Stimulant 12-Mo. Freq.	Meth. 12-Mo. Freq.				12-17	18-25	26+
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	26	1,2,5,7,10,15,16, 18,19,22,28,29	1. R1 2. R2 3. (R1+R2)*PY	1-4: 0 1-3: 0 1,2: 11	1-4: 0 1-3: 0 1,2: 9	1-4: 0 1-3: 0 1,2: 6
25	Past year	(Past year not past month)	Missing	Missing	0	1-3,5,7,8,15-19, 25,27,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
26	Past year	Lifetime	Missing	Missing	0	1,2,5,7,10,15,16, 18,19,22,29	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases

Meth. = methamphetamine.

NOTE: Users of stimulants included users of methamphetamine.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month stimulants use | lifetime stimulants use)
2. R2 = P(Past year but not past month stimulants use | lifetime stimulants use)
3. PY = P(Stimulants use on a given day in the past year | past year stimulants use)

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Pain Reliever Recency	Oxy. Recency	Pain Reliever 12-Mo. Freq.	Oxy. 12-Mo. Freq.				12-17	18-25	26+
1	(Past month)		Missing		57	1-5,23	1. PY	1-3,5: 31 1-3: 2	1-3,5: 14 1-3: 0 1,2: 1	1-3,5: 9
2	(Past year but not past month)		Missing		44	1-3,6,7	1. PY	1-3,5: 23 1-3: 2	1-3,5: 8	1-3,5: 9 1-3: 2
3	Past year				3	7,8	1. R1/(R1+R2)	1-3,5: 2	1-3,5: 1	No Cases
4	Past year		Missing		26	1-3,5,7,8,25	1. R1/(R1+R2) 2. PY	1-3,5: 1 1-3: 1 1,2: 4	1-3,5: 2 1-3: 1 1,2: 9	1-3,5: 3 1-3: 2 1,2: 3
5	Missing (lifetime use known)		Missing		313	1-3,5,7,25	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 26 1-3: 2 1,2: 145	1-3,5: 16 1-3: 2 1,2: 59	1-3,5: 6 1-3: 0 1,2: 57
6	(Past month)	(Past month)		Missing	2	4,15-20,24	1. PY	1-3,5: 1	No Cases	1-3,5: 0 1-3: 0 1,2: 1
7	(Past year not missing)	(Past year not past month)		Missing	1	8,14-20	1. PY	1-3,5: 0 1-3: 0 1,2: 1	No Cases	No Cases
8	(Past year not missing)	Past year			1	7,8	1. R1/(R1+R2)	No Cases	1-3,5: 1	No Cases
9	(Past year not missing)	Past year	Missing		0	1-3,7,8,23	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
10	(Past year not missing)	Past year		Missing	2	7,8,15-17,26	1. R1/(R1+R2) 2. PY	1-3,5: 2	No Cases	No Cases
11	(Past year not missing)	Past year	Missing	Missing	0	1-3,7,8,15-17, 23,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Pain Reliever Recency	Oxy. Recency	Pain Reliever 12-Mo. Freq.	Oxy. 12-Mo. Freq.				12-17	18-25	26+
12	(Past year not missing)	Missing (lifetime use known)		Missing	19	7,8,15-17	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 5	1-3,5: 0 1-3: 0 1,2: 11	1-3,5: 0 1-3: 0 1,2: 3
13	(Past month)	(Past month)	Missing	Missing	0	1-4,7,15-17,20, 23,24	1. PY	No Cases	No Cases	No Cases
14	(Past month)	(Past year not past month)	Missing	Missing	0	1-4,7,15-17,21, 23	1. PY	No Cases	No Cases	No Cases
15	(Past year not past month)	(Past year not past month)	Missing	Missing	2	1-3,6,7,15-17,21	1. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 1	No Cases
16	Past year	Past year			1	7,8,14,19,28	1. R1/(R1+R2)	1-3,5: 1	No Cases	No Cases
17	Past year	Past year	Missing		0	1-3,7,8,14,19, 25,28	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
18	Past year	Past year		Missing	10	7,8,14-17,19, 26,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 3	1-3,5: 0 1-3: 0 1,2: 6	1-3,5: 0 1-3: 0 1,2: 1
19	Past year	Past year	Missing	Missing	7	1-3,7,8,14-17, 19,25,26,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 1 1-3: 0 1,2: 4	1-3,5: 0 1-3: 0 1,2: 1
20	Past year	Missing (lifetime use known)		Missing	7	7,8,10,15-17, 19,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 1	1-3,5: 1 1-3: 0 1,2: 4	1-3,5: 0 1-3: 0 1,2: 1
21	Past year	Missing (lifetime use known)	Missing	Missing	3	1-3,7,8,10, 15-17,19,25,28	1. R1/(R1+R2) 2. PY	1-3,5: 0 1-3: 0 1,2: 2	1-3,5: 0 1-3: 0 1,2: 1	No Cases
22	(Past month)	Missing (lifetime use known)		Missing	0	4,7,10,15-17,19	1. R1 2. R2 3. (R1+R2)*PY	No Cases	No Cases	No Cases
23	(Past month)	Missing (lifetime use known)	Missing	Missing	3	1-4,7,10,15-17, 19,23	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 0 1-3: 0 1,2: 2	1-3,5: 0 1-3: 0 1,2: 1	No Cases

Table E.54 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Recency and Frequency (continued)

#	Missingness Pattern				Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Pain Reliever Recency	Oxy. Recency	Pain Reliever 12-Mo. Freq.	Oxy. 12-Mo. Freq.				12-17	18-25	26+
24	Missing (lifetime use known)	Missing (lifetime use known)	Missing	Missing	44	1-3,7,10,15-17, 19,22,28	1. R1 2. R2 3. (R1+R2)*PY	1-3,5: 1 1-3: 0 1,2: 16	1-3,5: 0 1-3: 0 1,2: 19	1-3,5: 1 1-3: 0 1,2: 7
25	Past year	(Past year not past month)	Missing	Missing	0	1-3,7,8,14-17, 19,25	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases
26	Past year	Lifetime	Missing	Missing	0	1-3,6-8,10, 15-17,19,25	1. R1/(R1+R2) 2. PY	No Cases	No Cases	No Cases

Oxy. = OxyContin.

NOTE: Users of pain relievers included users of OxyContin.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month pain relievers use | lifetime pain relievers use)
2. R2 = P(Past year but not past month pain relievers use | lifetime pain relievers use)
3. PY = P(Pain relievers use on a given day in the past year | past year pain relievers use)

Table E.55 Logical Constraints for Core-Plus-Noncore Stimulants Recency

Constraint #	Logical Constraint
LogC8	Donor must be a past month or past year but not past month stimulant user (stimulant recency = 1 or 2)
LogC10	Donor must be a past month, past year but not past month, or lifetime but not past year methamphetamine user (methamphetamine recency = 1, 2, or 3)
LogC14	Donor must be a past month or past year but not past month methamphetamine user (methamphetamine recency = 1 or 2)
LogC22	Donor must be a past month, past year but not past month, or lifetime but not past year stimulant user (stimulant recency = 1, 2, or 3)
LogC27	If recipient is not a lifetime user of any type of stimulant except for methamphetamine, then donor must not have used stimulants more recently than recipient has used methamphetamines
LogC28	If recipient is not a lifetime user of any type of stimulants except for methamphetamines, then donor must not have used stimulants more recently than donor has used methamphetamines

Table E.56 Likeness Constraints for Core-Plus-Noncore Stimulants Recency

Constraint #	Likeness Constraint
LikC1	Donor's stimulant recency must match recipient's stimulant recency (when nonmissing)[1]
LikC2	State rank of donor = State rank of recipient
LikC3	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC4	Donor's methamphetamine recency agrees with recipient's methamphetamine recency (when nonmissing)

[1] Although this constraint also is used as a logical constraint for some missingness patterns, it is included for clarity.

Table E.57 Constraints and Portion of the Predictive Mean Vector for Core-Plus-Noncore Stimulants Recency

#	Missingness Pattern		Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
	Stimulant Recency	Meth. Recency				12-17	18-25	26+
1	Past year		3	8,27	1. R1/(R1+R2)	No Cases	1-4: 2	1-4: 0 1-3: 0 1,2: 1
2	Missing (lifetime use known)		55	22,27	1. R1 2. R2	1-4: 1 1-3: 3 1,2: 19	1-4: 19 1-3: 0 1,2: 9	1-4: 0 1-3: 0 1,2: 4
3	(Past month)	Past year	2	14	1. R1/(R1+R2)	No Cases	1-4: 1 1-3: 0 1,2: 1	No Cases
4	(Past year not missing)	Missing (lifetime use known)	2	10	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 1	1-4: 0 1-3: 0 1,2: 1	No Cases
5	Past year	Past year	5	8,14,28	1. R1/(R1+R2)	1-4: 1 1-3: 0 1,2: 1	1-4: 0 1-3: 0 1,2: 2	1-4: 1
6	Past year	Missing (lifetime use known)	3	8,10,28	1. R1/(R1+R2)	1-4: 1 1-3: 0 1,2: 1	1-4: 1	No Cases
7	Missing (lifetime use known)	Missing (lifetime use known)	42	10,22,28	1. R1 2. R2	1-4: 0 1-3: 0 1,2: 10	1-4: 2 1-3: 0 1,2: 7	1-4: 2 1-3: 0 1,2: 21

Meth. = methamphetamine.

NOTE: Users of stimulants included users of methamphetamine.

[1] The predictive mean vector components are defined by the following:
1. R1 = P(Past month stimulants use | lifetime stimulants use)
2. R2 = P(Past year but not past month stimulants use | lifetime stimulants use)
3. PY = P(Stimulants use on a given day in the past year | past year stimulants use)

E.3.3 Age at First Use

Table E.58 Logical Constraints for Age at First Use, Univariate Assignment

Constraint #	Logical Constraint
LogC1	Donor's age at first drug use must be less than or equal to recipient's age
LogC2	If recipient was not a past year drug user, then donor cannot have age at first drug use equal to recipient's age
LogC3	If recipient was not a past 3 years drug user, then donor cannot have age at first drug use equal to recipient's age, recipient's age minus 1, or recipient's age minus 2
LogC4	If recipient's 12-month drug frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first drug use equal to recipient's age
LogC5	If recipient's 30-day drug frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first drug use equal to recipient's age
LogC6	If recipient's most recent birthday was within the past 30 days, and recipient's drug recency was past year but not past month, then donor cannot have age at first drug use equal to recipient's age
LogC7	If recipient's drug recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month drug frequency was less than 30, then donor cannot have age at first drug use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first drug use must be earlier than recipient's most recent birthday.)
LogC8	Donor's age at first cigarette use must be less than or equal to recipient's age at first daily cigarette use (if existing)
LogC9	Donor's age at first daily cigarette use must be greater than or equal to recipient's imputation-revised age at first cigarette use

Table E.59 Likeness Constraints for Age at First Use, Univariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	If recipient was a past year drug user, then donor was a past year drug user; if recipient was a past 3 years but not past year drug user, then donor was the same; if recipient was a lifetime but not past 3 years drug user, then donor was the same
LikC4	Age of donor = Age of recipient
LikC5	Age of donor must be greater than or equal to age of recipient
LikC6	If recipient was a past year drug user, then donor was a past year drug user; if recipient was not a past year drug user, then donor was the same
LikC7	If recipient was a past 3 years but not past year drug user, then donor was not a past year drug user; if recipient was a lifetime but not past 3 years drug user, then donor was the same
LikC8	If recipient was not a past year drug user, then donor was the same

Table E.60 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	444	1-3,5,6,8	1. PrAFU	1-4: 212 2-4: 2 3,4: 1	1-4: 128	1-4: 92 2-4: 6 3,4: 2 4,7: 0 5,7: 0 7: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first use

Table E.61 Constraints and Portion of the Predictive Mean Vector for Cigarette Age at First Daily Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	117	1-3,5,6,9	1. PrAFU	1-4: 10 2-4: 0 3,4: 3 4,7: 0 5,7: 0 7: 0 None: 2	1-4: 19 2-4: 2	1-4: 69 2-4: 5 3,4: 5 4,7: 0 5,7: 2

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigarette age at first daily use

Table E.62 Constraints and Portion of the Predictive Mean Vector for Cigar Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	439	1-3,5,6	1. PrAFU	1-4: 115 2-4: 8 3,4: 1	1-4: 136	1-4: 161 2-4: 11 3,4: 6 4,7: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cigar age at first use

Table E.63 Constraints and Portion of the Predictive Mean Vector for Alcohol Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	645	1,2,4,6,7	1. PrAFU	1,2,4,6: 306	1,2,4,6: 125 2,4,6: 1	1,2,4,6: 206 2,4,6: 6 4,6: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted alcohol age at first use

Table E.64 Constraints and Portion of the Predictive Mean Vector for Inhalants Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	293	1,2,4,6,7	1. PrAFU	1,2,4,6: 205 2,4,6: 0 4,6: 2	1,2,4,6: 44 2,4,6: 1	1,2,4,6: 35 2,4,6: 3 4,6: 3

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted inhalants age at first use

Table E.65 Constraints and Portion of the Predictive Mean Vector for Marijuana Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	210	1,2,4,6,7	1. PrAFU	1,2,4,6: 90	1,2,4,6: 55	1,2,4,6: 62 2,4,6: 2 4,6: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted marijuana age at first use

Table E.66 Constraints and Portion of the Predictive Mean Vector for Tranquilizers Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	173	1,2,4,6,7	1. PrAFU	1,2,4,6: 54 2,4,6: 2	1,2,4,6: 53	1,2,4,6: 38 2,4,6: 10 4,6: 10 4,8: 2 5,8: 3 8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted tranquilizers age at first use

Table E.67 Constraints and Portion of the Predictive Mean Vector for Sedatives Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	66	1,2,4,6,7	1. PrAFU	1,2,4,6: 17 2,4,6: 3 4,6: 4	1,2,4,6: 12 2,4,6: 1 4,6: 3	1,2,4,6: 10 2,4,6: 8 4,6: 7 4,8: 0 5,8: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted sedatives age at first use

Table E.68 Constraints and Portion of the Predictive Mean Vector for Heroin Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	12	1,2,4,6,7	1. PrAFU	1,2,4,6: 3	1,2,4,6: 5	1,2,4,6: 0 2,4,6: 4

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted heroin age at first use

Table E.69 Logical Constraints for Age at First Use, Multivariate Assignment

Constraint #	Logical Constraint
LogC1	For each parent or child drug for which recipient was missing age at first use, donor's age at first use must be less than or equal to recipient's age
LogC2	For each parent or child drug for which recipient was missing age at first use, if recipient was not a past year user, then donor cannot have age at first use equal to recipient's age
LogC3	For each child drug for which recipient was missing age at first use, if recipient was not a past 3 years user, then donor cannot have age at first use equal to recipient's age, recipient's age minus 1, or recipient's age minus 2
LogC4	If recipient was missing the age at first use for the parent drug, and if recipient's parent drug 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have parent drug age at first use equal to recipient's age
LogC5	For each child drug for which recipient was missing age at first use, if recipient's 12-month frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first use equal to recipient's age
LogC6	For each child drug for which recipient was missing age at first use, if recipient's 30-day frequency was greater than the number of days since recipient's most recent birthday, then donor cannot have age at first use equal to recipient's age
LogC7	For each parent or child drug for which recipient was missing age at first use, if recipient's most recent birthday was within the past 30 days, and recipient's recency was past year but not past month, then donor cannot have age at first use equal to recipient's age
LogC8	If recipient was missing the age at first use for the parent drug, and recipient's parent drug recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's parent drug 12-month frequency was less than 30, then donor cannot have parent drug age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)
LogC9	For each child drug for which recipient was missing age at first use, if recipient's recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and recipient's 12-month frequency was less than 30, then donor cannot have age at first use equal to recipient's age. (These recipients are missed by LogC3 and LogC4, but the idea is the same: the date of first use must be earlier than recipient's most recent birthday.)
LogC10	For each child drug for which recipient was missing age at first use, donor must have age at first use greater than or equal to recipient's parent drug age at first use (if existing)
LogC11	If recipient was missing the age at first use for the parent drug, donor must have parent drug age at first use less than or equal to the minimum of recipient's child drug(s) age(s) at first use (if existing)
LogC12	For each child drug for which recipient was missing age at first use, donor must be a lifetime user
LogC13	If recipient was not a lifetime user of "other" pain relievers/stimulants, then donor must have age at first use for the parent drug equal to age at first use for the child drug
LogC14	If recipient was not a lifetime user of "other" hallucinogens, then donor must not have values that would cause the minimum of the child drug(s) age(s) at first use to be greater than the overall hallucinogens age at first use for recipient. For example, if donor was not a lifetime user of "other" hallucinogens and has HALLAGE = 17, LSDAGE = missing, PCPAGE = 19, and ECSAGE = missing, then donor must have LSDAGE = 17 and/or ECSAGE = 17.

Table E.69 Logical Constraints for Age at First Use, Multivariate Assignment (continued)

Constraint #	Logical Constraint
LogC15	For each child drug for which recipient was missing age at first use, if recipient's recency was past year but not past month, and the difference between the number of days since recipient's most recent birthday and the parent drug 12-month frequency was less than 30, then donor cannot have age at first use equal to recipient's age. (This constraint technically eliminates some potential donors who should be eligible, but it would be difficult to write a similar constraint that eliminates exactly the right donors from the neighborhood. For example, consider an 18-year-old respondent with missing LSDAGE whose birthday was 40 days ago, has IRLSDRC = 2, and has IRHALFY = 15. There is no question in the questionnaire for LSD 12-month frequency, so unless respondent did not use any hallucinogens other than LSD, the LSD 12-month frequency is unknown. The "conservative" constraint is to eliminate anyone from the neighborhood with IRLSDAGE = 18.)

Table E.70 Likeness Constraints for Age at First Use, Multivariate Assignment

Constraint #	Likeness Constraint
LikC1	State rank of donor = State rank of recipient
LikC2	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC3	For each child drug for which recipient was missing age at first use, if recipient was a past year user, then donor was a past year user; if recipient was a past 3 years but not past year user, then donor was the same; if recipient was a lifetime but not past 3 years user, then donor was the same
LikC4	Age of donor = Age of recipient
LikC5	Age of donor must be greater than or equal to age of recipient
LikC6	For the parent drug, if recipient was a past year user, then donor was a past year user; if recipient was a lifetime but not past year user, then donor was the same
LikC7	For each child drug, if recipient was a past year user, then donor was either a past year user or a lifetime nonuser; if recipient was a lifetime but not past year user, then donor was either a lifetime but not past year user or a lifetime nonuser[1]
LikC8	Donor agrees with recipient with respect to lifetime use for child drug(s)
LikC9	For the parent drug, if recipient was a lifetime but not past year user, then donor was a lifetime but not past year user
LikC10	For each child drug, if recipient was a lifetime but not past year user, then donor was not a past year user[1]
LikC11	Donor was at least as old as recipient, but no more than 20 years older than recipient
LikC12	Donor was no more than 20 years older than recipient
LikC13	For each child drug for which recipient was missing age at first use, if recipient was a past 3 years but not past year child drug user, then donor was not a past year child drug user; if recipient was a lifetime but not past 3 years child drug user, then donor was the same

[1] These constraints were intended to match recency of use for child drugs. They were not applied exactly as intended, because lifetime nonusers were not correctly considered. However, the practical impact was small because child drug nonusers could not be donors for a missing child drug age at first use.

Table E.71 Constraints and Portion of the Predictive Mean Vector for Smokeless Tobacco Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	316	1-3,6,7,12	1. PrAFU	1-4: 95 2-4: 11 3,4: 7 4,13: 0 5,13: 1	1-4: 129 2-4: 2	1-4: 49 2-4: 10 3,4: 9 4,13: 3

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted smokeless tobacco age at first use

Table E.72 Constraints and Portion of the Predictive Mean Vector for Cocaine Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	75	1,2,4,5,7-12	1. PrAFU	1,2,4,6-8: 7 2,4,6-8: 1 2,4,6,7: 0 4,6,7: 0 4,9,10: 0 9-11: 1	1,2,4,6-8: 19 2,4,6-8: 0 2,4,6,7: 0 4,6,7: 2 4,9,10: 1	1,2,4,6-8: 33 2,4,6-8: 4 2,4,6,7: 1 4,6,7: 6

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted cocaine age at first use

Table E.73 Constraints and Portion of the Predictive Mean Vector for Hallucinogens Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	207	1,2,4,7,8, 10-12,14,15	1. PrAFU	1,2,4,6-8: 34 2,4,6-8: 13 2,4,6,7: 6 4,6,7: 8 4,9,10: 0 9-11: 2	1,2,4,6-8: 55 2,4,6-8: 4 2,4,6,7: 1 4,6,7: 5 4,9,10: 3	1,2,4,6-8: 41 2,4,6-8: 13 2,4,6,7: 11 4,6,7: 9 4,9,10: 0 9-11: 1 9,10,12: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted hallucinogens age at first use

Table E.74 Constraints and Portion of the Predictive Mean Vector for Pain Relievers Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	646	1,2,4,5,7-13	1. PrAFU	1,2,4,6-8: 255 2,4,6-8: 9 2,4,6,7: 0 4,6,7: 1	1,2,4,6-8: 187 2,4,6-8: 1	1,2,4,6-8: 145 2,4,6-8: 21 2,4,6,7: 2 4,6,7: 21 4,9,10: 3 9-11: 1

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted pain relievers age at first use

Table E.75 Constraints and Portion of the Predictive Mean Vector for Stimulants Age at First Use

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26+
1	Missing	124	1,2,4,5,7-13	1. PrAFU	1,2,4,6-8: 26 2,4,6-8: 7 2,4,6,7: 0 4,6,7: 5 4,9,10: 0 9-11: 5	1,2,4,6-8: 32 2,4,6-8: 4 2,4,6,7: 1	1,2,4,6-8: 22 2,4,6-8: 9 2,4,6,7: 1 4,6,7: 7 4,9,10: 1 9-11: 4

[1] The predictive mean vector components are defined by the following:
1. PrAFU = Predicted stimulants age at first use

E.4 Household Composition (Roster) Variables

Tables E.76 through E.97 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the eight household composition (roster) variables: number of rostered persons, number of children younger than 18, number of persons aged 65 or older, indicator of whether the respondent has family members in household, number of respondent's family members in the household (both including and excluding foster relationships), and number of respondent's family members younger than 18 in the household (both including and excluding foster relationships).

Table E.76 Likeness Constraints for Number of Rostered Persons

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means

Table E.77 Constraints and Portion of the Predictive Mean Vector for Number of Rostered Persons

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	42	None	1. C1	1: 8	1: 15	1: 18	1: 1

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.78 Logical Constraints for Number of Children Younger than 18

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.79 Likeness Constraints for Number of Children Younger than 18

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.80 Constraints and Portion of the Predictive Mean Vector for Number of Children Younger than 18

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	225	1	1. C1	1,2: 80 2: 0 None: 1	1,2: 78	1,2: 60 2: 0 None: 1	1,2: 5

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.81 Logical Constraints for Number of Persons Aged 65 or Older

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE, IRKID17, and nonmissing ages in the roster

Table E.82 Likeness Constraints for Number of Persons Aged 65 or Older

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.83 Constraints and Portion of the Predictive Mean Vector for Number of Persons Aged 65 or Older

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	377	1	1. C1	1,2: 197	1,2: 123	1,2: 52 1: 1	1,2: 4

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.84 Likeness Constraints for Indicator of Whether the Respondent Has Family Members in Household

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	Donor's predicted means must be within 20 percent of recipient's predicted means
LikC3	IRKID17 of donor = IRKID17 of recipient
LikC4	If recipient was married, then donor was married; otherwise, if recipient was not currently married, then donor was not currently married

Table E.85 Constraints and Portion of the Predictive Mean Vector for Indicator of Whether the Respondent Has Family Members in Household

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	50	None	1. F1	1-4: 12	1-4: 17 2-4: 1	1-4: 20	No Cases

[1] The predictive mean vector components are defined by the following:
1. F1 = P(No other family members in the household)

Table E.86 Logical Constraints for Number of Respondent's Family Members in Household (including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRHHSIZE and nonmissing ages in the roster

Table E.87 Likeness Constraints for Number of Respondent's Family Members in Household (including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient

Table E.88 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	75	1	1. C1	1-3: 26	1-3: 24	1-3: 25	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.89 Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (including Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFMLYSZ and nonmissing ages in the roster

Table E.90 Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (including Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient

Table E.91 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (including Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	91	1	1. C1	1-4: 41 2-4: 0 3,4: 0 4: 0 None: 1	1-4: 18 2-4: 1	1-4: 29 2-4: 0 3,4: 1	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.92 Logical Constraints for Number of Respondent's Family Members in Household (excluding Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFMLYSZ and nonmissing ages in the roster

Table E.93 Likeness Constraints for Number of Respondent's Family Members in Household (excluding Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient

Table E.94 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household (excluding Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	68	1	1. C1	1-4: 24	1-4: 21	1-4: 23	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

Table E.95 Logical Constraints for Number of Respondent's Family Members in Household Younger than 18 (excluding Foster Relationships)

Constraint #	Logical Constraint
LogC1	Lower and upper bounds were restricted based on IRFAMSZE and nonmissing ages in the roster

Table E.96 Likeness Constraints for Number of Respondent's Family Members in Household Younger than 18 (excluding Foster Relationships)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means
LikC2	IRKID17 of donor = IRKID17 of recipient
LikC3	IRHHSIZE of donor = IRHHSIZE of recipient
LikC4	IRFMLYSZ of donor = IRFMLYSZ of recipient
LikC5	IRFAMSZE of donor = IRFAMSZE of recipient
LikC6	IRKDFMLY of donor = IRKDFMLY of recipient

Table E.97 Constraints and Portion of the Predictive Mean Vector for Number of Respondent's Family Members in Household Younger than 18 (excluding Foster Relationships)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	83	1	1. C1	1-6: 39 2-6: 0 3-6: 0 4-6: 0 5,6: 0 6: 1	1-6: 16	1-6: 27	No Cases

[1] The predictive mean vector components are defined by the following:
1. C1 = Predicted count from Poisson regression model

E.5 Income Variables

Tables E.98 through E.103 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the income variables.

E.5.1 Binary Variable Phase

Five of the binary income variables were directly related to a respondent's socioeconomic status. Hence, if a recipient required imputation for one or more of these five variables (i.e., welfare payments, welfare services, food stamps, binary income, and months on welfare), but had information on at least one of these variables, the donors were restricted so that donors and recipients had the same values for these nonmissing variables. These five variables are referred to as "welfare-correlated variables."

There were a large number of missingness patterns for the source-of-income variables because they are imputed simultaneously in a set. A respondent could be missing any combination of the seven source-of-income variables. Because the constraints and predictive mean vectors can be described easily, only one row in Table E.100 is used to summarize the many multiple variable missingness patterns.

Table E.98 Logical Constraints for Source of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient[1]
LogC2	Recipient is missing months on welfare. If recipient is also known to have received either welfare payments or welfare services, then donor must also have received welfare payments or welfare services. (This prevents donor from giving a skip code to a recipient who should have a value from 1 to 12.)

[1] This is only a logical constraint when family binary total income is missing.

Table E.99 Likeness Constraints for Source of Income

Constraint #	Likeness Constraint
LikC1	Age of donor = Age of recipient
LikC2	If any of the welfare-correlated edited binary income variables (welfare payments, welfare services, food stamps, binary total income, and months on welfare) were missing, then donor must match recipient on all nonmissing welfare-correlated edited binary income variables
LikC3	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing family variables
LikC4	If recipient is missing months on welfare, then donor must match recipient with respect to personal welfare payments (if nonmissing) and welfare services (if nonmissing)
LikC5	If recipient is missing social security, then donor must match recipient with respect to whether there are adults aged 65 or older in the household
LikC6	If recipient is missing welfare payments and/or welfare services, then donor must match recipient with respect to whether there are children younger than 18 in the household
LikC7	If recipient is missing wages, then donor must match recipient with respect to whether there are adults aged 18 to 64 in the household
LikC8	Age of donor must be within 5 years of age of recipient
LikC9	If recipient is missing binary income at the family level but not at the personal level, then donor must match recipient with respect to binary income at the personal level
LikC10	If recipient is not missing binary income at the personal level, then donor must match recipient with respect to binary income at the personal level

Table E.100 Constraints and Portion of the Predictive Mean Vector for Source of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing welfare months, receiving family payment and/or family service	179	1,2	1. WMS, and probabilities associated with other missing elements	1-7: 26 1,2,4-7: 43 2,4-8: 1	1-7: 28 1,2,4-7: 48 2,4-8: 1	1-7: 1 1,2,4-7: 21 2,4-8: 8	1-7: 0 1,2,4-7: 0 2,4-8: 1 2,4-7: 1
2	Missing welfare months, not receiving welfare payments, missing welfare services	97	1,2	1. SVC*WMS, SVC, and probabilities associated with other missing elements	1-7: 24 1,2,4-7: 19	1-7: 13 1,2,4-7: 30	1-7: 2 1,2,4-7: 7	1-7: 0 1,2,4-7: 2
3	Missing welfare months, missing welfare payments, not receiving welfare services	214	1,2	1. PMT*WMS, PMT, and probabilities associated with other missing elements	1-7: 59 1,2,4-7: 39	1-7: 20 1,2,4-7: 62	1-7: 1 1,2,4-7: 30	1-7: 0 1,2,4-7: 3
4	Missing welfare months, missing welfare payments, missing welfare services	287	1,2	1. [1-(1-PMT)(1-SVC)]*WMS, PMT, SVC, and probabilities associated with other missing elements	1-7: 49 1,2,4-7: 133	1-7: 1 1,2,4-7: 75	1-7: 2 1,2,4-7: 21	1-7: 0 1,2,4-7: 5 2,4-8: 1

Table E.100 Constraints and Portion of the Predictive Mean Vector for Source of Income (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
5	All other missingness patterns with number of months on welfare nonmissing[2]	3,444	1	1. Probabilities associated with missing elements	1-7: 926 1,2,5-7: 82	1-7: 1,382 1,2,5-7: 164 2,5-8: 2	1-7: 522 1,2,5-7: 156 2,5-8: 5	1-7: 110 1,2,5-7: 87 2,5-8: 8

[1] The predictive mean vector components are defined by the following:
1. PMT = P(Family in household received income from welfare payments)
2. SVC = P(Family in household received income from other welfare services)
3. WMS = P(Family in household received any welfare on a given month in the past year | family received any welfare in the past year)

[2] There were many other missingness patterns for source of income because a respondent could be missing any combination of the 10 source-of-income variables. Because the constraints and predictive mean vectors can be described easily, all of these other missingness patterns are represented by a single row in the table.

E.5.2 Specific Category Phase

Table E.101 Logical Constraints for Finer Categories of Income

Constraint #	Logical Constraint
LogC1	IRFAMSKP of donor = IRFAMSKP of recipient
LogC2	Personal (PINC2) and family (FINC2) income of donor must be consistent with personal and family income of recipient

Table E.102 Likeness Constraints for Finer Categories of Income

Constraint #	Likeness Constraint
LikC1	Donor's predicted mean must be within 5 percent of recipient's predicted mean
LikC2	PINC2 of donor = PINC2 of recipient, if nonmissing
LikC3	FINC2 of donor = FINC2 of recipient, if nonmissing
LikC4	IRPINC1 of donor = IRPINC1 of recipient
LikC5	IRFINC1 of donor = IRFINC1 of recipient

Table E.103 Constraints and Portion of the Predictive Mean Vector for Finer Categories of Income

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Completely missing	7,114	1,2	1. Xβ	1-5: 2,188 2-5: 3	1-5: 2,852 2-5: 3	1-5: 1,591 2-5: 1	1-5: 470 2-5: 6

[1] For more details, see Section 9.3.2.2.

E.6 Health Insurance Variables

Tables E.104 through E.109 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the health insurance variables. Tables E.104 through E.107 are for variables created using the "Constituent Variables Method," and Tables E.108 and E.109 are for variables created using the "Old Method." See Chapter 10 for an explanation of the two methods.

In several instances, variable names are used without description for the purposes of brevity (see Chapter 10 for details). For the health insurance imputations, matches between donors and recipients were attempted on the nonmissing values of the variables CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN. These variables are the edited indicators of whether the respondent received health insurance from Medicaid/State health insurance programs for children, Medicare, CHAMPUS, or private health insurance, respectively. These were the base variables used in the creation of the imputation-revised variables (IRMCDCHP, IRMEDICR, IRCHMPUS, IRPRVHLT, and IROTHHLT). In addition to the edited health insurance variables, other variables, which were used as likeness constraints, are identified in the tables only by their variable names. These include SERVICE (an indicator of whether the respondent had ever been

in the military), GOVTPROG (an indicator of whether the respondent's family participated in government public assistance programs), INCOME (a four-level categorical family-income variable, with levels of less than $20,000, $20,000 to less than $50,000, $50,000 to less than $75,000, and $75,000 or more), IRFAMIN1 (a two-level family income variable, with levels of less than $20,000 and $20,000 or more), and IRFAMSOC (an indicator of whether the respondent's family in the household received income from social security).

E.6.1 Constituent Variables Method

For the MPMN, the likeness constraints, which were applied to the variables, differed between missingness patterns, and sometimes the constraints differed between age groups within the same missingness pattern.

Table E.104 Likeness Constraints for Health Insurance, Constituent Variables Method (MPMN)

Constraint #	Likeness Constraint
LikC1	Donor's predicted means must be within 5 percent of recipient's predicted means for all missing constituent variables (CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN)
LikC2	GOVTPROG of donor = GOVTPROG of recipient
LikC3	IRFAMSOC of donor = IRFAMSOC of recipient
LikC4	SERVICE of donor = SERVICE of recipient (if nonmissing)
LikC5	INCOME of donor = INCOME of recipient
LikC6	IRFAMIN1 of donor = IRFAMIN1 of recipient
LikC7	Donor must match recipient for all nonmissing constituent variables
LikC8	If recipient is between 18 and 64 years old and has a nonmissing value for edited work status (JBSTATR), then (1) if recipient has no job due to disability (JBSTATR = 14), then donor must have no job due to disability (2) if recipient has JBSTATR not equal to 14, then donor must have JBSTATR not equal to 14

MPMN = multivariate predictive mean neighborhood.

Table E.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (MPMN)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Only missing CAIDCHIP	327	None	1. HI1	1,2,7: 160 1,2: 2	1,2,7: 117 1,2: 2	1,2,7: 30	1,2,7: 16
2	Only missing MEDICARE	68	None	1. HI2	1,3,7: 18 1,3: 1	1,3,7,8: 36	1,3,7,8: 11	1,3,7: 2
3	Missing CAIDCHIP and MEDICARE	53	None	1. HI1 2. HI2	1-3,7: 8 1-3: 3 1: 1	1,2,7,8: 26 1,2,8: 2 1,8: 1 8: 2	1,2,7,8: 7 1,2,8: 0 1,8: 1 8: 1	1,2,7: 1
4	Only missing CHAMPUS	79	None	1. HI3	1,4,7: 30 1,4: 1 4: 1	1,4,7: 33	1,4,7: 11	1,4,7: 3
5	Missing CAIDCHIP and CHAMPUS	26	None	1. HI1 2. HI3	1,2,4,7: 13 1,2,4: 1 1,4: 1	1,2,4,7: 11	No Cases	No Cases
6	Missing MEDICARE and CHAMPUS	4	None	1. HI2 2. HI3	1,3,4,7: 1 1,3,4: 1	1,4,7,8: 1 1,4,8: 1	No Cases	No Cases
7	Missing CAIDCHIP, MEDICARE, and CHAMPUS	17	None	1. HI1 2. HI2 3. HI3	1-4,7: 4 1-4: 3 1,4: 0 4: 2	1,2,4,7,8: 1 1,2,4,8: 1 1,4,8: 0 4,8: 3	1,2,4,7,8: 0 1,2,4,8: 0 1,4,8: 0 4,8: 3	No Cases
8	Only missing PRVHLTIN	241	None	1. HI4	1,5,7: 106 1,5: 0 1,6: 1	1,5,7: 106	1,5,7: 17	1,6,7: 11
9	Missing CAIDCHIP and PRVHLTIN	90	None	1. HI1 2. HI4	1,2,5,7: 59 1,2,5: 2 1,6: 1 1: 0 None: 1	1,2,5,7: 19 1,2,5: 2 1,6: 3	1,2,5,7: 2	1,2,6,7: 0 1,2,6: 0 1: 0 None: 1
10	Missing MEDICARE and PRVHLTIN	11	None	1. HI2 2. HI4	1,3,5,7: 3 1,3,5: 0 1,6: 1 1: 1 None: 3	1,5,7,8: 3	No Cases	No Cases

Table E.105 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (MPMN) (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
11	Missing CAIDCHIP, MEDICARE, and PRVHLTIN	18	None	1. HI1 2. HI2 3. HI4	1-3,5,7: 3 1-3,5: 0 1,6: 0 1: 2 None: 2	1,2,5,7,8: 2 1,2,5,8: 0 1,6,8: 1 1,8: 1 8: 6	1,2,5,7,8: 0 1,2,5,8: 0 1,6,8: 0 1,8: 0 8: 1	No Cases
12	Missing CHAMPUS and PRVHLTIN	25	None	1. HI3 2. HI4	1,4,5,7: 8 1,4,5: 1 1,4,6: 0 1,4: 0 4: 3	1,4,5,7: 4 1,4,5: 0 1,4,6: 1 1,4: 1 4: 5	1,4,5,7: 1 1,4,5: 1	1,4,6,7: 0 1,4,6: 0 1,4: 0 4: 0
13	Missing CAIDCHIP, CHAMPUS, and PRVHLTIN	41	None	1. HI1 2. HI3 3. HI4	1,2,4,5,7: 12 1,2,4,5: 0 1,4,6: 0 1,4: 1 4: 16	1,2,4,5,7: 3 1,2,4,5: 0 1,4,6: 1 1,4: 2 4: 6	No Cases	No Cases
14	Missing MEDICARE, CHAMPUS, and PRVHLTIN	5	None	1. HI2 2. HI3 3. HI4	1,3-5,7: 0 1,3-5: 0 1,4,6: 0 1,4: 0 4: 3	1,4,5,7,8: 0 1,4,5,8: 0 1,4,6,8: 0 1,4,8: 0 4,8: 2	No Cases	1,4,6,7: 0 1,4,6: 0 1,4: 0 4: 0
15	Missing CAIDCHIP, MEDICARE, CHAMPUS, and PRVHLTIN	88	None	1. HI1 2. HI2 3. HI3 4. HI4	1-5: 20 1,4,6: 0 1,4: 0 4: 37	1,2,4,5,8: 3 1,4,6,8: 0 1,4,8: 0 4,8: 19	1,2,4,5,8: 2 1,4,6,8: 0 1,4,8: 0 4,8: 6	1,2,4,6: 0 1,4: 0 4: 1

MPMN = multivariate predictive mean neighborhood.

[1] The predictive mean vector components are defined by the following:
1. HI1 = P(Respondent has any other health insurance)
2. HI2 = P(Respondent has Medicare)
3. HI3 = P(Respondent has CHAMPUS)
4. HI4 = P(Respondent has private health insurance)

Table E.106 Constraints for Health Insurance, Constituent Variables Method (UPMN), Any Other Health Insurance

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is between 26 and 64 years old, donor must also be between 26 and 64 years old; if recipient is aged 65 years or older, donor must also be aged 65 years or older

UPMN = univariate predictive mean neighborhood.

Table E.107 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Constituent Variables Method (UPMN), Any Other Health Insurance

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group		
					12-17	18-25	26 or Older
1	Missing ANYOTHER	208	None	1. HI1	1: 49	1: 130	1,2: 29

UPMN = univariate predictive mean neighborhood.

[1] The predictive mean vector components are defined by the following:
1. HI1 = P(Respondent has any other health insurance)

E.6.2 Old Method

Table E.108 Likeness Constraints for Health Insurance, Old Method

Constraint #	Likeness Constraint
LikC1	Age of donor = Age of recipient
LikC2	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC3	Age of donor must be within 5 years of age of recipient
LikC4	Donor must not have received private health insurance (PINSUR = 0)
LikC5	Donor must not have received overall health insurance by the 1999 definition (INSUR = 0)
LikC6	Donor must have received overall health insurance by the 2001 definition (INSUR3 = 1)
LikC7	Donor must have received overall health insurance by the 1999 definition (INSUR = 1)

Table E.109 Constraints and Portion of the Predictive Mean Vector for Health Insurance, Old Method

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases, by Age Group			
					12-17	18-25	26-64	65 or Older
1	Missing INSUR3, no for INSUR, and no for PINSUR	65	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4,5: 38	1,2,4,5: 23	1,2,4,5: 4	No Cases
2	Yes for INSUR3, missing INSUR, and no for PINSUR	22	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4,6: 18	1,2,4,6: 4	No Cases	No Cases
3	Missing INSUR3, missing INSUR, and no for PINSUR	157	None	1. (OVR*(1-PRV))/ (1-OVR*PRV)	1,2,4: 62	1,2,4: 71	1,2,4: 24	No Cases
4	Yes for INSUR3, missing INSUR, missing PINSUR	25	None	1. OVR 2. OVR*PRV	1,2,6: 21 1,6: 1	1,2,6: 3	No Cases	No Cases
5	Missing INSUR3, missing INSUR, missing PINSUR	392	None	1. OVR 2. OVR*PRV	1,2: 208 1: 5	1,2: 151 1: 1	1,2: 23 1: 3	1,2: 0 1: 1
6	Yes for INSUR3, yes for INSUR, and missing for PINSUR	102	None	1. PRV	1,2,7: 51	1,2,7: 35	1,2,7: 4	1,2,7: 11 1,7: 0 3,7: 1

[1] The predictive mean vector components are defined by the following:
1. OVR = P(Respondent received health insurance, 2001 definition)
2. PRV = P(Respondent received private health insurance | respondent received health insurance, 2001 definition)

E.7 Roster Pair Variables

Tables E.110 through E.129 present information on the missingness patterns, constraints, and predictive mean vectors applied during the imputation procedures for the pair variables. Tables E.110 through E.113 correspond to the first stage of imputation, where the pair relationship variable (IRPRREL) is created. Tables E.114 through E.121 are for the second stage, imputation of the multiplicity variables. Finally, Tables E.122 through E.129 correspond to the third stage, where the household counts are imputed.

There are a few instances where variable names are used without description for the purposes of brevity. In these cases, the variables are defined in a table at the beginning of the section in which they are used.

E.7.1 Stage One: Pair Relationship

Table E.110 Variables Used in Constraints for Stage One, Pair Relationship

Variable	Value	Constraint
MARIT1	1	At least one pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
MARIT2	2	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	1	One pair member has a marital status of "Married" that was not imputed or logically assigned
	0	Neither pair member has a marital status of "Married," "Widowed," or "Divorced or Separated" that was not imputed or logically assigned
MARIT3	5	Both pair members have a marital status of "Married" that was not imputed or logically assigned
	4	One pair member has a marital status of "Married," and the other has a marital status of "Widowed" or "Divorced or Separated," neither of which was imputed or logically assigned
	3	One pair member has a marital status of "Married," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	2	Both pair members have a marital status of "Widowed" or "Divorced or Separated" that was not imputed or logically assigned
	1	One pair member has a marital status of "Widowed" or "Divorced or Separated," and the other either has a marital status of "Never Married" or was younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned
	0	Both pair members either have a marital status of "Never Married" or were younger than 15 and therefore had a legitimate skip for marital status, neither of which was imputed or logically assigned

Table E.111 Logical Constraints for Stage One, Pair Relationship

Constraint #	Logical Constraint
LogC10	If recipient is a spouse-spouse pair, with or without children, then donor must be a spouse-spouse pair either with or without children
LogC15	If recipient could be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 12 to 14 or some other clearly identifiable pair that is not of interest
LogC16	If recipient could be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 15 to 17 or some other clearly identifiable pair that is not of interest
LogC17	If recipient could be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 18 to 20 or some other clearly identifiable pair that is not of interest
LogC18	If recipient could be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest, then donor must be either a parent-child pair where the child is aged 21 or older or some other clearly identifiable pair that is not of interest
LogC20	If recipient could be either a sibling-sibling pair where one sibling is aged 12 to 17 and the other is aged 18 to 25 or some other clearly identifiable pair that is not of interest, then donor must be either a sibling-sibling (12-17/18-25) pair or some other clearly identifiable pair that is not of interest
LogC21	If recipient could be either a sibling-sibling pair where the siblings are not in either age range of interest or some other clearly identifiable pair that is not of interest, then donor must be a sibling-sibling pair where the siblings are not in an age range of interest or some other clearly identifiable pair that is not of interest
LogC22	If recipient could be either a spouse-spouse pair with children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with children or some other clearly identifiable pair that is not of interest
LogC23	If recipient could be either a spouse-spouse pair without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair without children or some other clearly identifiable pair that is not of interest
LogC24	If recipient could be a spouse-spouse pair with or without children or another clearly identifiable pair that is not of interest, then donor must be a spouse-spouse pair with or without children or some other clearly identifiable pair that is not of interest
LogC25	If recipient could be either a grandparent-grandchild pair or another clearly identifiable pair that is not of interest, then donor must be a grandparent-grandchild pair or some other clearly identifiable pair that is not of interest

Table E.112 Likeness Constraints for Stage One, Pair Relationship

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient is not a possible parent-child pair, and recipient has at least one parent in the household, then donor must also have at least one parent in the household
LikC3	If recipient is not a possible parent-child pair, and recipient has no parents in the household, then donor must also not have any parents in the household
LikC4	If the number of children aged 0 to 18[1] is nonmissing for both recipient and donor, then the two values must be equal
LikC5	If the recipient pair members are of the same gender, then the donor pair members must be of the same gender. If the recipient pair members are of a different gender, then the donor pair members must also be of a different gender.
LikC6	The age of the donor's younger pair member must be the same age as the recipient's younger pair member
LikC61	If donor is a spouse-spouse pair, with or without children, then donor's younger pair member must be the same age as recipient's younger pair member
LikC62	If both donor and recipient have pair members of different ages, then donor's younger pair member must be the same age as recipient's younger pair member
LikC7	The younger pair member of both donor and recipient must fall within the same age group: 21-25, 26-34, 35-49, or 50 or older
LikC8	The older pair member of both donor and recipient must fall within the same age group: 26-34, 35-49, or 50 or older
LikC9	If recipient pair members are in the same age group (21-25, 26-34, 35-49, or 50 or older), then donor cannot be a parent-child pair where the child is aged 21 or older
LikC10	Neither of the donor's pair members can have had an imputed marital status
LikC101	If donor is a sibling-sibling (12-14/15-17) pair, then neither of donor's pair members can have had an imputed marital status
LikC11	If neither of the recipient's pair members' marital status was imputed or if recipient has MARIT1 = 1, then donor must have the same value for MARIT1 as recipient
LikC12	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT2 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC121	If donor is a sibling-sibling (12-14/15-17) pair, then apply Likeness Constraint 12
LikC13	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value for MARIT3 as recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT2 = 1, then donor must have had MARIT2 = 1 or 2.
LikC14	If neither of the recipient's pair members had an imputed marital status, then donor must have the same value of MARIT3 as the recipient. However, if one of the recipient's pair members had an imputed marital status, and recipient had MARIT1 = 1, then donor must also have had MARIT1 = 1.

[1] For age group pairs 3 and 4, this constraint is on the number of children aged 0 to 11.

Table E.113 Constraints and Portion of the Predictive Mean Vector for Stage One, Pair Relationship

#	Age Group Pair[1]	Total Number of Cases	Logical Constraints	Predictive Mean Vector[2]	Likeness Constraints: Number of Cases
1	One pair member is aged 12 to 14, the other is aged 15 to 17	10	None	1. SIB	1-3: 10
2	One pair member is aged 12 to 14, the other is aged 18 to 25	11	15,20	1. SIB	1-3,10,11: 10 2,3,10,11: 1
3	Both pair members are aged 15 to 17	4	10,16,21-25	1. SIB	1-5,62,101,121: 4
4	One pair member is aged 15 to 17, the other is aged 18 to 25	16	10,16,20-25	1. SIB	1-6,10,13: 10 2-6,10,13: 6
5	One pair member is aged 18 to 20, the other is aged 18 to 25	30	10,17,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,13: 8 2-5,10,13: 22
6	Both pair members are aged 21 to 25	44	10,18,21-24	1. SPOUSE1 2. SPOUSE2	1-5,10,12: 16 2-5,10,12: 28
7	One pair member is aged 12 to 14, the other is aged 26 or older	5	15,21,25	1. PC	1-3,8,10,11: 2 2,3,8,10,11: 3
8	One pair member is aged 15 to 17, the other is aged 26 or older	6	16,21-25	1. PC	1-5,61,8,10,14: 2 2-5,61,8,10,14: 4
9	One pair member is aged 18 to 20, the other is aged 26 or older	14	10,17,21-25	1. PC	1-5,8,10,14: 6 2-5,8,10,14: 7 2-5,8,10,11: 1
10	One pair member is aged 21 or older, the other is aged 26 or older	73	10,18,21-25	1. SPOUSE1 2. SPOUSE2	1-5,7-10,12: 7 2-5,7-10,12: 64 2-5,9,10,12: 2
11	Both pair members are aged 12 to 17	6	None	1. SIB	1-3: 5 2: 1

[1] Because there was only one variable to be imputed, PAIRREL, there was only one missingness pattern. However, as the predictive mean vector and constraints vary by age group, the age groups are presented as missingness patterns in this table, and the column heading has been changed accordingly.

[2] The predictive mean vector components are defined by the following:
1. SIB = P(the pair relationship is sibling-sibling)
2. SPOUSE1 = P(the pair relationship is spouse-spouse, with children)
3. SPOUSE2 = P(the pair relationship is spouse-spouse, without children)
4. PC = P(the pair relationship is parent-child)

E.7.2 Stage Two: Multiplicity Counts

Table E.114 Logical Constraints for Stage Two, Multiplicity Counts

Constraint #	Logical Constraint
LogC2	If recipient is a parent-child pair and if recipient's household count(s) in the appropriate age group(s) (12-14, 15-17, and/or 12-20) is nonmissing, then donor's and recipient's count(s) must match. Also, donor's parent-focus multiplicity count(s) must fall within recipient's bounds.
LogC3	Donor's sibling-sibling (12-14/15-17) older focus multiplicity count must fall within recipient's bounds
LogC4	Donor's sibling-sibling (12-14/15-17) younger focus multiplicity count must fall within recipient's bounds
LogC5	Donor's sibling-sibling (12-17/18-25) older focus multiplicity count must fall within recipient's bounds
LogC6	Donor's sibling-sibling (12-17/18-25) younger focus multiplicity count must fall within recipient's bounds

Table E.115 Likeness Constraints for Stage Two, Multiplicity Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	Donor's pair relationship must match recipient's pair relationship
LikC3	Donor and recipient must have the same household size
LikC10	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's value must match
LikC11	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's value must match
LikC12	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's value must match
LikC13	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's value must match
LikC14	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's value must match
LikC15	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's value must match
LikC16	If recipient's count of household members aged 26 or older is nonmissing, then donor's value must match
LikC17	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's value must match
LikC18	If recipient's count of household members aged 50 or older is nonmissing, then donor's value must match

Table E.116 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Child Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Completely missing	106	None	1. PRED1	1-3,10,12,15,17,18: 50 2,3,10,12,15,17,18: 46 2,3,15,17,18: 7 2,16: 3

[1] PRED1 is the predicted mean for the number of parents of a child aged 12 to 20 who is a member of a parent-child pair.

Table E.117 Constraints and Portion of the Predictive Mean Vector for Parent-Child, Parent Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Completely missing	0	2	1. PRED1	No Cases

[1] PRED1 is the predicted mean for the number of children aged 12 to 20 belonging to a parent who is a member of a parent-child pair.

Table E.118 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	2	3	1. PRED1	1,3,10,11,12: 1 3,10,11,12: 0 3,11: 1

[1] PRED1 is the predicted mean for the number of siblings aged 12 to 14 for a respondent aged 15 to 17 who is a member of a sibling-sibling pair.

Table E.119 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-14/15-17), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	2	4	1. PRED1	1,3,10,11,12: 0 3,10,11,12: 2

[1] PRED1 is the predicted mean for the number of siblings aged 15 to 17 for a respondent aged 12 to 14 who is a member of a sibling-sibling pair.

Table E.120 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Older Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	5	5	1. PRED1	1,3,10,12,14: 5

[1] PRED1 is the predicted mean for the number of siblings aged 12 to 17 for a respondent aged 18 to 25 who is a member of a sibling-sibling pair.

Table E.121 Constraints and Portion of the Predictive Mean Vector for Sibling-Sibling (12-17/18-25), Younger Focus Multiplicity Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases
1	Missing	5	6	1. PRED1	1,3,10,12,14: 2 3,10,12,14: 0 3,14: 2 14: 1

[1] PRED1 is the predicted mean for the number of siblings aged 18 to 25 for a respondent aged 12 to 17 who is a member of a sibling-sibling pair.

E.7.3 Stage Three: Household Counts

Table E.122 Variables Used in Constraints for Stage Three, Household Counts

Variable	Value	Constraint
Males XXXX[1]	0	No household members in the specified age group are male
	1	All household members in the specified age group are male
	2	Some household members in the specified age group are male
	3	There are no household members in the specified age group

[1] XXXX denotes the 4-digit age range (e.g., 1217 denotes the age range 12 to 17).

Table E.123 Logical Constraints for Stage Three, Household Counts

Constraint #	Logical Constraint
LogC1	Donor's count of older siblings aged 15 to 17 must fall within recipient's bounds
LogC2	Donor's count of older siblings aged 15 to 17 cannot be larger than recipient's nonmissing count of household members aged 15 to 17
LogC3	Donor's count of older siblings aged 18 to 25 must fall within recipient's bounds
LogC4	Donor's count of older siblings aged 18 to 25 cannot be larger than recipient's nonmissing count of household members aged 18 to 25
LogC5	Donor's count of spouse-spouse pairs must fall within recipient's bounds
LogC6	Donor's count of spouse-spouse pairs with children must fall within recipient's bounds
LogC7	Donor's parent-child (12-14) parent count must be greater than 0
LogC8	Donor's parent-child (12-17) parent count must be greater than 0
LogC9	Donor's parent-child (12-20) parent count must be greater than 0
LogC10	Donor's parent-child (12-14) child count must be greater than 0
LogC11	Donor's parent-child (12-17) child count must be greater than 0
LogC12	Donor's parent-child (12-20) child count must be greater than 0
LogC13	Donor's parent-child (12-14) child count must fall within recipient's bounds
LogC14	Donor's parent-child (12-17) child count must fall within recipient's bounds
LogC15	Donor's parent-child (12-20) child count must fall within recipient's bounds
LogC16	Donor's parent-child (12-14) parent count must fall within recipient's bounds
LogC17	Donor's parent-child (12-17) parent count must fall within recipient's bounds
LogC18	Donor's parent-child (12-20) parent count must fall within recipient's bounds

Table E.124 Likeness Constraints for Stage Three, Household Counts

Constraint #	Likeness Constraint
LikC1	Donor's predicted means each must be within 5 percent of recipient's predicted means
LikC2	If recipient lives in a multi-family home, then donor must also live in a multi-family home
LikC3	Donor's household size must be the same as recipient's household size
LikC4	If recipient's count of household members aged 0 to 11 is nonmissing, then donor's count must match
LikC5	If recipient's count of household members aged 12 to 14 is nonmissing, then donor's count must match
LikC6	If recipient's count of household members aged 12 to 17 is nonmissing, then donor's count must match
LikC7	If recipient's count of household members aged 12 to 20 is nonmissing, then donor's count must match
LikC8	If recipient's count of household members aged 15 to 17 is nonmissing, then donor's count must match
LikC9	If recipient's count of household members aged 15 or older is nonmissing, then donor's count must match
LikC10	If recipient's count of household members aged 18 to 25 is nonmissing, then donor's count must match
LikC11	If recipient's count of household members aged 26 to 34 is nonmissing, then donor's count must match
LikC12	If recipient's count of household members aged 26 to 49 is nonmissing, then donor's count must match
LikC13	If recipient's count of household members aged 35 to 49 is nonmissing, then donor's count must match
LikC14	If recipient's count of household members aged 50 or older is nonmissing, then donor's count must match
LikC15	Donor's and recipient's counts of household members aged 0 to 17 must both be 0, 1, or greater than 1
LikC16	Donor's and recipient's counts of household members aged 0 to 17 must either both be 0 or both be positive
LikC17	Donor's and recipient's counts of household members aged 12 to 14 must either both be 0 or both be positive
LikC18	Donor's and recipient's counts of household members aged 12 to 17 must either both be 0 or both be positive
LikC19	Donor's and recipient's counts of household members aged 15 to 17 must either both be 0 or both be positive
LikC20	Donor's and recipient's counts of household members aged 18 to 25 must either both be 0 or both be positive
LikC21	If recipient's value of MALES15P is nonmissing, then donor's value of MALES15P must match
LikC22	If recipient's count of household members aged 15 or older is positive, then donor and recipient must have the same count of household members aged 15 or older and the same values for MALES15P

Table E.124 Likeness Constraints for Stage Three, Household Counts (continued)

Constraint #	Likeness Constraint
LikC23	If recipient's value of MALES1825 is nonmissing, then donor's value of MALES1825 must match
LikC24	If recipient's value of MALES1834 is nonmissing, then donor's value of MALES1834 must match
LikC25	If recipient's value of MALES2634 is nonmissing, then donor's value of MALES2634 must match
LikC26	If recipient's value of MALES2649 is nonmissing, then donor's value of MALES2649 must match
LikC27	If recipient's value of MALES2649 is nonmissing and not equal to 3, then donor's value of MALES2649 must match
LikC28	If recipient's value of MALES3549 is nonmissing, then donor's value of MALES3549 must match
LikC29	If recipient's value of MALES50P is nonmissing, then donor's value of MALES50P must match
LikC30	If recipient could possibly have a parent-child (12-14) parent count of 0, then donor's and recipient's counts of household members aged 12 to 14 must match
LikC31	If recipient could possibly have a parent-child (12-17) parent count of 0, then donor's and recipient's counts of household members aged 12 to 17 must match
LikC32	If recipient could possibly have a parent-child (12-20) parent count of 0, then donor's and recipient's counts of household members aged 12 to 20 must match
LikC33	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC34	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC35	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 34, 35 to 49, and 50 or older must match
LikC36	If recipient could possibly have a parent-child (12-14) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC37	If recipient could possibly have a parent-child (12-17) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC38	If recipient could possibly have a parent-child (12-20) child count of 0, then donor's and recipient's counts of household members aged 26 to 49 must match
LikC39	If recipient's count of household members aged 12 to 14 is nonmissing, then donor and recipient must have the same screener count of household members aged 12 to 14
LikC40	If recipient's count of household members aged 12 to 17 is nonmissing, then donor and recipient must have the same screener count of household members aged 12 to 17
LikC41	If recipient's count of household members aged 15 to 17 is nonmissing, then donor and recipient must have the same screener count of household members aged 15 to 17
LikC42	If recipient's count of household members aged 18 to 25 is nonmissing, then donor and recipient must have the same screener count of household members aged 18 to 25
LikC43	Donor and recipient must have parent-child (12-14) parent counts that are both 0 or both positive
LikC44	Donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive
LikC45	If recipient's counts of household members aged 12 to 14 and 15 to 17 are both nonmissing, then donor and recipient must have parent-child (12-14 and 12-17) parent counts that are both 0 or both positive
LikC46	If recipient's counts of household members aged 12 to 17 and 18 to 25 are both nonmissing, then donor and recipient must have parent-child (12-17) parent counts that are both 0 or both positive

Table E.125 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-14/15-17), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	34	1,2	1. H1	1,3-5,8,39,41,43,44: 8 3-5,8,39,41,43,44: 5 5,8,39,41,43,44: 3	1,3-5,8,39,41,43,44: 7 3-5,8,39,41,43,44: 9 5,8,39,41,43,44: 2

[1] H1 is the predicted mean count of older siblings aged 15 to 17.

Table E.126 Constraints and Portion of the Predictive Mean Vector for Household Sibling Count (12-17/18-25), Older Focus

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	52	3,4	1. H2	1,3,4,6,10,40,42,44: 10 3,4,6,10,40,42,44: 11 6,10,40,42,44: 8	1,3,4,6,10,40,42,44: 9 3,4,6,10,40,42,44: 7 6,10,40,42,44: 7

[1] H2 is the predicted mean count of older siblings aged 18 to 25.

Table E.127 Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	68	5	1. SPOUSE1 2. SPOUSE2	1-3,9-11,13-15,21,23,25,28,29: 7 2,3,9-11,13-15,21,23,25,28,29: 25 2,15,22,23,25,28,29: 10 2,22,24,26,29: 8	1-3,9-11,13-15,21,23,25,28,29: 10 2,3,9-11,13-15,21,23,25,28,29: 4 2,15,22,23,25,28,29: 1 2,22,24,26,29: 3

[1] The predictive mean vector components are defined by the following:
1. SPOUSE1 = P(the number of spouse-spouse pairs in the household is 0)
2. SPOUSE2 = P(the number of spouse-spouse pairs in the household is 1)

Table E.128 Constraints and Portion of the Predictive Mean Vector for Household Spouse-Spouse with Children Count

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing	36	6	1. SPOUSE3	1-3,9-11,13-15,21,23,25,28,29: 3 2,3,9-11,13-15,21,23,25,28,29: 13 2,15,22,23,25,28,29: 8 2,16,22,24,26,29: 4 2,16,22,27: 0 2,16,22,27*: 2	1-3,9-11,13-15,21,23,25,28,29: 6

*A donor could not be found for these cases. Imputed values were randomly assigned within the bounds created during the editing process.

[1] The predictive mean vector components are defined by the following:
1. SPOUSE3 = P(there is at least one spouse-spouse pair with children in the household)

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
1	Missing PPCP1220	6	9,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,32: 1 2-4,10,11,13,14,32: 3 2,4,10,11,13,14,32: 0 2,32: 1	1-4,10,11,13,14,32: 0 2-4,10,11,13,14,32: 1
2	Missing PPCP1217	4	8,17	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31: 2 2-4,10,11,13,14,31: 2	No Cases
3	Missing PPCP1217 and PPCP1220	23	8,9,17,18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,31,32: 4 2-4,10,11,13,14,31,32: 18 2,4,10,11,13,14,31,32: 1	No Cases
4	Missing PPCP1214	0	7,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
5	Missing PPCP1214 and PPCP1220	0	7,9,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
6	Missing PPCP1214 and PPCP1217	9	7,8,16,17	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30,31: 3 2-4,10,11,13,14,30,31: 6	No Cases
7	Missing PPCP1214, PPCP1217, and PPCP1220	47	7-9,16-18	1. PC1 2. PC2 3. H3	1-4,10,11,13,14,30-32: 6 2-4,10,11,13,14,30-32: 35 2,4,10,11,13,14,30-32: 1 2,32: 4	1-4,10,11,13,14,30-32: 1
8	Missing PPCC1220	19	12,15	1. PC1 2. PC2 3. H3	1-4,7,10,35: 1 2-4,7,10,35: 8 2,4,7,10,35: 2 2,7,38: 3	1-4,7,10,35: 3 2-4,7,10,35: 2
9	Missing PPCC1220 and PPCP1220	13	15,18	1. PC1 2. PC2 3. H3	1-4,7,10,11,13,14,32,35: 1 2-4,7,10,11,13,14,32,35: 5 2,4,7,10,11,13,14,32,35: 0 2,32: 4	1-4,7,10,11,13,14,32,35: 1 2-4,7,10,11,13,14,32,35: 1 2,4,7,10,11,13,14,32,35: 0 2,32: 1
10	Missing PPCC1220 and PPCP1217	0	8,12,15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
11	Missing PPCC1220, PPCP1217, and PPCP1220	0	8,15,17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
12	Missing PPCC1220 and PPCP1214	0	7,12,15,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
13	Missing PPCC1220, PPCP1214, and PPCP1220	0	7,15,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
14	Missing PPCC1220, PPCP1214, and PPCP1217	0	7,8,12,15-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
15	Missing PPCC1220, PPCP1214, PPCP1217, and PPCP1220	1	7,8,15-18	1. PC1 2. PC2 3. H3	1-4,7,10,11,13,14,30-32,35: 0 2-4,7,10,11,13,14,30-32,35: 0 2,4,7,10,11,13,14,30-32,35: 0 2,32: 1	No Cases

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
16	Missing PPCC1217	2	11,14	1. PC1 2. PC2 3. H3	1-4,6,10,34: 0 2-4,6,10,34: 2	No Cases
17	Missing PPCC1217 and PPCP1220	0	9,11,14,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
18	Missing PPCC1217 and PPCP1217	1	14,17	1. PC1 2. PC2 3. H3	1-4,6,10,11,13,14,31,34: 1	No Cases
19	Missing PPCC1217, PPCP1217, and PPCP1220	0	9,14,17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
20	Missing PPCC1217 and PPCP1214	0	7,11,14,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
21	Missing PPCC1217, PPCP1214, and PPCP1220	0	7,9,11,14,16, 18	1. PC1 2. PC2 3. H3	No Cases	No Cases
22	Missing PPCC1217, PPCP1214, and PPCP1217	0	7,14,16,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
23	Missing PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	7,9,14,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
24	Missing PPCC1217 and PPCC1220	14	11,12,14,15	1. PC1 2. PC2 3. H3	1-4,6,7,10,34,35: 2 2-4,6,7,10,34,35: 12	No Cases
25	Missing PPCC1217, PPCC1220, and PPCP1220	0	11,14,15,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
26	Missing PPCC1217, PPCC1220, and PPCP1217	0	12,14,15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
27	Missing PPCC1217, PPCC1220, PPCP1217, and PPCP1220	18	14,15,17,18	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,31,32,34,35: 4 2-4,6,7,10,11,13,14,31,32,34,35: 8 2,4,6,7,10,11,13,14,31,32,34,35: 0 2,32: 1	1-4,6,7,10,11,13,14,31,32,34,35: 3 2-4,6,7,10,11,13,14,31,32,34,35: 2
28	Missing PPCC1217, PPCC1220, and PPCP1214	0	7,11,12,14-16	1. PC1 2. PC2 3. H3	No Cases	No Cases
29	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	7,11,14-16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
30	Missing PPCC1217, PPCC1220, PPCP1214, and PPCP1217	0	7,12,14-17	1. PC1 2. PC2 3. H3	No Cases	No Cases
31	Missing PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	1	7,14-18	1. PC1 2. PC2 3. H3	1-4,6,7,10,11,13,14,30-32,34,35: 0 2-4,6,7,10,11,13,14,30-32,34,35: 0 2,4,6,7,10,11,13,14,30-32,34,35: 0 2,32: 1	No Cases
32	Missing PPCC1214	0	10,13	1. PC1 2. PC2 3. H3	No Cases	No Cases
33	Missing PPCC1214 and PPCP1220	0	9,10,13,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
34	Missing PPCC1214 and PPCP1217	0	8,10,13,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
35	Missing PPCC1214, PPCP1217, and PPCP1220	0	8-10,13,17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
36	Missing PPCC1214 and PPCP1214	0	13,16	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
37	Missing PPCC1214, PPCP1214, and PPCP1220	0	9,13,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
38	Missing PPCP1214 and PPCP1217	0	8,13,16,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
39	Missing PPCP1214, PPCP1217, and PPCP1220	0	8,9,13,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
40	Missing PPCC1214 and PPCC1220	0	10,12,13,15	1. PC1 2. PC2 3. H3	No Cases	No Cases
41	Missing PPCC1214, PPCC1220, and PPCP1220	0	10,13,15,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
42	Missing PPCC1214, PPCC1220, and PPCP1217	0	8,10,12,13, 15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
43	Missing PPCC1214, PPCC1220, PPCP1217, and PPCP1220	0	8,10,13,15, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
44	Missing PPCC1214, PPCC1220, and PPCP1214	0	12,13,15,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
45	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1220	0	13,15,16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
46	Missing PPCC1214, PPCC1220, PPCP1214, and PPCP1217	0	8,12,13,15-17	1. PC1 2. PC2 3. H3	No Cases	No Cases

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
47	Missing PPCC1214, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	0	8,13,15-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
48	Missing PPCC1214 and PPCC1217	0	10,11,13,14	1. PC1 2. PC2 3. H3	No Cases	No Cases
49	Missing PPCC1214, PPCC1217, and PPCP1220	0	9-11,13,14,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
50	Missing PPCC1214, PPCC1217, and PPCP1217	0	10,13,14,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
51	Missing PPCC1214, PPCC1217, PPCP1217, and PPCP1220	0	9,10,13,14, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
52	Missing PPCC1214, PPCC1217, and PPCP1214	0	11,13,14,16	1. PC1 2. PC2 3. H3	No Cases	No Cases
53	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1220	0	9,11,13,14, 16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
54	Missing PPCC1214, PPCC1217, PPCP1214, and PPCP1217	0	13,14,16,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
55	Missing PPCC1214, PPCC1217, PPCP1214, PPCP1217, and PPCP1220	0	9,13,14,16-18	1. PC1 2. PC2 3. H3	No Cases	No Cases
56	Missing PPCC1214, PPCC1217, and PPCC1220	4	10-15	1. PC1 2. PC2 3. H3	1-7,10,33-35: 0 2-7,10,33-35: 2	1-7,10,33-35: 0 2-7,10,33-35: 1 2,4-7,10,33-35: 1

Table E.129 Constraints and Portion of the Predictive Mean Vector for Household Parent-Child Counts (continued)

#	Missingness Pattern	Total Number of Cases	Logical Constraints	Predictive Mean Vector[1]	Likeness Constraints: Number of Cases	
					Responding Pairs	Non-Pair Members
57	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1220	0	10,11,13-15, 18	1. PC1 2. PC2 3. H3	No Cases	No Cases
58	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1217	0	10,12-15,17	1. PC1 2. PC2 3. H3	No Cases	No Cases
59	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1217, and PPCP1220	0	10,13-15, 17,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
60	Missing PPCC1214, PPCC1217, PPCC1220, and PPCP1214	2	11-16	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30,33-35: 0 2-7,10,11,13,14,30,33-35: 1 2,4-7,10,11,13,14,30,33-35: 0 2,30: 1	No Cases
61	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1220	0	11,13-16,18	1. PC1 2. PC2 3. H3	No Cases	No Cases
62	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, and PPCP1217	2	12-17	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30,31,33-35: 0 2-7,10,11,13,14,30,31,33-35: 0 2,4-7,10,11,13,14,30,31,33-35: 0 2,31: 2	No Cases
63	Missing PPCC1214, PPCC1217, PPCC1220, PPCP1214, PPCP1217, and PPCP1220	36	13-18	1. PC1 2. PC2 3. H3	1-7,10,11,13,14,30-35: 0 2-7,10,11,13,14,30-35: 18 2,4-7,10,11,13,14,30-35: 0 2,32: 10	1-7,10,11,13,14,30-35: 2 2-7,10,11,13,14,30-35: 6

[1] The predictive mean vector components are defined by the following:
1. H3 = Predicted mean count of children aged 12 to 20 with parents in the household
2. PC1 = P(there is one parent with at least one child aged 12 to 20 in the household)
3. PC2 = P(there are two parents with at least one child aged 12 to 20 in the household)

Appendix F: Quality Control Measures Used in the Imputation Procedures

Appendix F: Quality Control Measures Used in the Imputation Procedures

F.1 Introduction

For the 2012 National Survey on Drug Use and Health (NSDUH), a number of quality control (QC) measures were implemented for the imputation of demographic, drug use, income, health insurance, nicotine dependence, household composition (roster), and pair variables. These QC measures spanned the following three basic steps within the predictive mean neighborhood (PMN) imputation methodology: (1) response propensity adjustment, (2) prediction modeling, and (3) hot-deck imputation.

Checklists for the imputation of these variables were used during imputation processing for the 2012 survey and serve as formal documentation of the QC measures that were implemented. Most of these checks required both a technician check, where the individual who ran the computer program (technician) entered his or her name and the date the check was performed, and an independent reviewer check, where another person performed a check of the same items. The reviewer check ensured greater quality in the imputation procedures. Additionally, some checklist entries required the technician to document the procedures that were taken to run the programs, such as listing the variables that were dropped from the model in order to achieve model convergence.

The checks involved in each of the PMN steps are described in detail in the following sections. Variable-specific consistency checks were often created in addition to the checks of the PMN steps. However, they are not discussed in this appendix. The QC measures that were used in the editing process or the delivery of variables to other NSDUH teams also are not discussed.

F.2 Step 1: Response Propensity Adjustment

The first step of the PMN methodology, response propensity adjustment, adjusts the sampling weights for item nonresponse so that the item respondent weights that are used only during the imputation process are representative of the entire domain of interest. The item response propensity is modeled as a function of a predetermined set of covariates. The model can be thought of as a special case of the generalized exponential model (GEM)[1] developed for weighting procedures, in that imputations that are done at the item level are similar in nature to the weight adjustments made for entire units.

The following QC measures were conducted in Step 1:

- The output of the response propensity modeling program was checked for singularities. Any singularities that occurred were investigated, and the model was corrected by removing the correlated covariates from the model.

[1] The GEM macro, which was written in SAS/IML® software, was developed at RTI International for weighting procedures and is described in detail in Appendix A of Chen et al. (2014).

- The values of the convergence flags were checked to see whether the GEM converged.
- The current distance to root was checked to make sure it was close to 0, indicating that GEM converged.
- When variables were reduced from the original model, the remaining levels of variables were checked to ensure appropriateness. An example of this check was to determine whether the base variables or lower order terms were present when interactions or higher order terms existed (e.g., "age" and "age squared" must be in the model if "age cubed" is in the model).
- An indicator was calculated in the response propensity program that measured the maximum adjustment to the weights. In most cases, the adjusted weights resembled the original design weights. If the maximum adjustment was too high (usually greater than 3.00), this was likely due to an overspecified model, where the adjustment was not performing at an optimum level. Large maximum adjustments were investigated and corrected, if possible, by removing extraneous variables from the models so that any final weight adjustment applied to a respondent was within acceptable bounds.
- After the weights were adjusted, the ratio of the maximum adjusted weight to the mean adjusted weight (called the "mmratio") was computed to monitor the variation among the weights. Any mmratio value that was greater than 25 percent was noted in the response propensity program checklist.
- The unequal weighting effect (UWE) was checked before and after adjustment to ensure that there was no significant variance increase due to the nonresponse adjustment. The difference in the UWE after-adjustment value should be no more than 20 percent of the UWE before-adjustment value. The difference was fairly small in most cases, and any difference greater than 20 percent was investigated and corrected, if possible.
- When using the SAS procedure PROC MEANS to examine summary statistics, the weighted totals for the independent variables in the model were compared both before and after the adjustment. If these weighted totals were equal, the adjustment procedures worked properly.

F.3 Step 2: Prediction Modeling

The prediction modeling step, using the weights from Step 1, calculates predicted means, which are then used in the hot-deck step(s) to create neighborhoods and select donors. The dependent variable in the model is usually the variable, or some transformation of that variable, that is undergoing imputation. Each model is built using only those cases within the domain with complete responses for that item. Predicted means are then calculated for all of the domain members, whether or not they were item respondents, using the values for the covariates and the estimates for the regression coefficients.

For categorical outcome variables, logistic regression models are used for the prediction models. For continuous variables, linear regression models are fit. For count variables, Poisson

regression models are used. For response variables that are proportions (e.g., months on welfare), a logit transformation is applied to the proportion, and a linear regression model is utilized. For continuous nonnegative random variables (e.g., finer categories of income), a failure time model is used.

The models are discussed in detail in the imputation sections of Chapters 4 through 11 of this report. The following QC measures were conducted in Step 2:

- All models were checked for singularities and collinearities. Any singularities that occurred were investigated, and the model was corrected.
- For Poisson regression models, failure time models, and logistic models, the output was checked to ensure convergence had occurred.
- For logistic models, the log file was checked for "data warning" messages or other SUDAAN®-specific errors.[2] If there was a "data warning" message in the log, the SUDAAN model was determined to be unstable, and variables were removed to produce stability in the estimates. Similar to the response propensity model, if the main variable was dropped, its interaction variables also were dropped.
- When variables were reduced from the original model, the remaining levels of variables were checked to ensure appropriateness.
- Output was checked to verify that everything worked properly in the regression model.
- A check on the predicted means from the model was created to ensure that each respondent in the domain had a valid nonmissing predicted mean.

F.4 Step 3: Hot-Deck Imputation

The predicted means from Step 2 were used to determine the final assignments of imputed values in a hot-deck step. The goal of this step was to make donors and recipients as similar as possible. A neighborhood of potential donors was used, if possible, so that the donor selected was different each time the procedure was run. However, all potential donors in a neighborhood usually had very similar predicted means.

The QC measures in this step had two objectives: (1) ensure that the imputed values were consistent with preexisting nonmissing values, and (2) ensure that the imputed values were assigned as intended. Many variables had logical constraints that had to be adhered to in order to prevent logical inconsistencies. Specific QC checks were implemented to ensure this; however, only the more general checks that apply to all PMN imputations are listed in this appendix. For more detail on logical constraints, refer to Appendix E. The following QC measures were conducted in Step 3:

- Unusual imputed values were noted. If the imputed value was equivalent to one of the standard NSDUH missing value codes, this signaled a failure to obtain a donor, and measures were required (e.g., likeness constraints were loosened) to

[2] Details can be found in the *SUDAAN® Language Manual, Release 10.0* (RTI International, 2008).

revise the programs so that a donor could be found. If the imputed value was otherwise unusual, the imputation process was examined to ensure that no errors occurred.

- The number of cases that were imputed within various levels of restrictiveness of the likeness constraints (as determined by the variable SMALLFLG) was noted.[3]
- The frequency of the variable "WORKED" was checked to ensure that no values were equal to 0. Values greater than 0 signified that the imputation procedure was able to find a donor for all missing cases.
- The distribution of edited variables was compared with the distribution of imputed variables to make sure that each imputed value was within the appropriate range corresponding to the value of the edited variable.
- The imputed values were crossed with the imputation indicators to ensure that the indicators were created correctly.
- After imputation was implemented, the distribution of values for nonrespondents was checked against the distribution of values for all respondents to ensure the similarity of these two subgroups.
- It was necessary to ensure that everyone to whom the variable did not apply received a skip code for the final imputed variable. For example, all those in the 12-to-14 age group should have a "Not Applicable (N/A)" value of 99 for the imputation-revised marital status variable IRMARIT.
- Any missing values were noted. This occurred when the program was unsuccessful in assigning a valid imputed value.
- Any cases where the imputed value was not consistent with preexisting nonmissing values were noted.
- It was necessary to ensure that any restrictions on the final imputed value for a given nonrespondent were honored. For example, some respondents were known to be employed, but either full-time or part-time employment status was not known. Checks were conducted to ensure that these respondents had either full-time or part-time status assigned to the employment status variable (EMPSTAT4), but not unemployed or other statuses.
- When multiple missingness patterns existed, each pattern was treated separately, and the distribution of imputed values within each missingness pattern was investigated. For example, if it was known that a respondent was a past year user, both past month and past year users should have been included among the imputed values, not just past month users.

[3] Refer to Appendix E for more details about likeness constraints and the "SMALLFLG" variable.

Appendix G: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

Appendix G: Interviewer Explanations for Overrides to Consistency Checks in Household Roster

G.1 Introduction

In the household roster for the 2012 National Survey on Drug Use and Health (NSDUH), the interviewer was supposed to enter a roster of the respondent's entire household, which included age, gender, and the relationship to the respondent. It was not uncommon for the interviewer to enter a relationship code, age, or gender that was illogical, based on the age and gender of the respondent given in the core part of the questionnaire. Before the computer-assisted interviewing (CAI) instrument was implemented in the 1999 NSDUH, such responses would have been flagged at the data processing stage.[1] However, beginning with the 2000 survey and in every survey year since then, consistency checks have been added to the CAI instrument that allowed the interviewer, if needed, to correct the error while giving the interview. Details about these consistency checks are presented in Chapter 8.

In general, two types of consistency checks were implemented in the 2012 survey. The first type compared the entry in the roster with previously entered questionnaire information, specifically the respondent's age (CURNTAGE) and gender. The second type checked for internal consistency within the household roster. In some cases, a consistency check would be triggered even though the response was legitimate. This occurred if CURNTAGE was considered incorrect, for example, or in extremely rare family situations such as when a stepmother was younger than her stepson. With the exception of the check against the previously entered respondent's gender, the interviewer could override the consistency check and explain why the response given was correct. In some cases, the interviewer was correct in overriding the consistency check. In others, however, it was clear that the interviewer misunderstood how the roster should have been put together and that the override to the consistency check was not legitimate.

This appendix summarizes the explanations given by interviewers for consistency check overrides in the household roster. It is divided into two parts: consistency check overrides involving CURNTAGE and those involving internal consistency checks.

G.2 Override Comments from Interviewers: Comparisons with CURNTAGE

When an interviewer entered the respondent's roster entry (the "self" entry), if the age did not match the age previously entered in the questionnaire, a consistency check was triggered. The comparison was between the roster age for the "self" and CURNTAGE, which was the value of the final questionnaire-edited age (AGE) stored by Blaise.[2] Explanations for consistency check overrides for the variable CURNTAGE are provided in Table G.1. Because CURNTAGE

[1] Because the age and gender of the respondent given in the core part of the questionnaire were not allowed to change, the relationship code and sometimes the age of the roster member were set to bad data.

[2] The Blaise program is the computer program within the CAI instrument that was used to direct the respondent and interviewer through the questionnaire.

had the potential to change constantly throughout the questionnaire, no final variable with this name was created. However, in most cases, the value of CURNTAGE when the roster commenced was equivalent to NEWAGE, the value of CURNTAGE after the drug modules had been completed. In theory, NEWAGE was not always equivalent to AGE, the derivation of which is described in Chapter 4.

In the 2002 survey, the explanations provided in Table G.1 were not reviewed when determining AGE, nor were they reviewed when determining the final value of the age for the "self" entry in the roster. However, beginning with the 2003 survey, these explanations have been carefully reviewed. In rare cases, the final value of AGE was set to the age of the self in the questionnaire roster (the "roster age"), based on these explanations as well as other evidence, even if it disagreed with the age as it would have been calculated in prior survey years. Details about this process are provided in Chapter 4.

Even in cases where the explanation seemed clear that CURNTAGE was incorrect, the value of AGE was not always set to the roster age. In most cases, this was because the difference between CURNTAGE and the roster age was 1 year or less. A difference of 1 year was tolerated, because legitimate differences could result from a birthday occurring in the time between the drug modules and the roster.[3] In other situations, the value of CURNTAGE was incorrect, but the original questionnaire-edited age was correct, so no change was necessary. In still other cases, not all the criteria that were necessary for changing the value of AGE to be equal to the roster age were met. Cases where the value of AGE was changed to roster age are denoted in the "Comment" column in bolded italics in Table G.1. Otherwise, the reason for not changing the value of AGE to roster age also is shown in this column. The "Respondent's Age in Roster of Other Pair Member" column indicates whether the roster of the other pair member, if it existed, supported CURNTAGE or the override age as the respondent's age.

G.3 Override Comments from Interviewers: Internal Consistency Check Overrides

Internal consistency checks were performed on the household roster for the 2012 NSDUH. Interviewer explanations for overrides to these internal consistency checks are provided in Table G.2. These explanations were evaluated individually to determine their legitimacy. Also provided in this table are the questionnaire-edited age of the respondent (AGE), the age and relationship to the respondent of the roster member in question, and, in the "Comment" column, an evaluation of whether the override was considered legitimate. If the override was considered legitimate, no edit was applied to the age or relationship code of the roster member. If the override was not considered legitimate, the override was overruled, and the relationship code (and sometimes the roster member's age) was set to bad data. In this instance, a brief indication of the probable true relationship of the roster member to the respondent is provided in the "Comment" column of the table.

[3] It was not uncommon for an interview to be conducted in more than one sitting. This could have occurred if either the respondent or the interviewer did not have enough time for the interview or otherwise could not complete the interview in a single sitting.

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
1	19	18	19	respondent was 18 on last bday; is now 19	Not in a pair	Diff. ≤ 1 year	19
2	25	24	25	this is the r	Not in a pair	Diff. ≤ 1 year	25
3	23	24	23	IR is 24	23	Diff. ≤ 1 year	23
4	17	16	17	respondent is 17 now. was 16 on last birthday	Not in a pair	Diff. ≤ 1 year	17
5	21	20	21	person # 2 is 21 years old	Not in a pair	Diff. ≤ 1 year	21
6	25	24	25	24 is her correct age	25	Diff. ≤ 1 year	25
7	24	23	23	IR is 24 as her birthday is on March; SR, aunt, said she was 23	Not in a pair	Diff. ≤ 1 year	24
8	15	14	15	respondent was 14 on last birthday which was in march	15	Diff. ≤ 1 year	15
9	20	19	19	the correct age is 19 in the interview	Not in a pair	***AGE was changed to equal roster age***	19
10	12	13	12	AGE WRONG	Not in a pair	Diff. ≤ 1 year	12
11	13	12	12	12	13	Diff. ≤ 1 year	13
12	13	12	13	R states that she is 12	13	Diff. ≤ 1 year	13
13	31	30	30	age is actually 32	Not in a pair	Diff. ≤ 1 year	31
14	12	11	12	respondent was 11 in january, 2012	12	Diff. ≤ 1 year	12
15	12	11	12	response by interview respondent	Not in a pair	Diff. ≤ 1 year	12
16	28	27	27	Respodent states that she wrongly entered her age as 28 in an alcohol question	28	Diff. ≤ 1 year	28
17	47	48	48	hit supress in error	47	Diff. ≤ 1 year	47
18	17	16	16	Sister gave him as 16 not 17	Not in a pair	Diff. ≤ 1 year	17
19	15	14	13	accidently entered	Not in a pair	Diff. ≤ 1 year	15
20	13	12	13	'last' birthday 12 years old	Not in a pair	Diff. ≤ 1 year	13
21	26	25	25	R stated that she is 25 years old	26	Diff. ≤ 1 year	26
22	20	21	21	r is 20 not 21 she is r5623612a sister is 19	Not in a pair	Diff. ≤ 1 year	20
23	37	36	37	37	37	Diff. ≤ 1 year	37
24	52	51	51	R is 51	52	Diff. ≤ 1 year	52

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
25	26	25	25	ir is 25	Not in a pair	Diff. ≤ 1 year	26
26	14	13	14	Because respondent was 13 on last birthday	Not in a pair	Diff. ≤ 1 year	14
27	26	25	24	this is the information that she is giving me	Not in a pair	Diff. ≤ 1 year	26
28	19	18	18	ir bdate feb 28 1994, gave 1993 on first question	19	Diff. ≤ 1 year	19
29	27	28	28	data entry error , r is 28	Not in a pair	Diff. ≤ 1 year	27
30	46	45	46	R had trouble understanding question re. his age	46	Diff. ≤ 1 year	46
31	16	17	17	SR stated that R is 17 but he is going to be 17 in Aug 1	16	Diff. ≤ 1 year	16
32	58	57	57	she is firm about been 57 even when computer calculated 58	Not in a pair	Diff. ≤ 1 year	58
33	14	13	14	this is correct interviewer	14	Diff. ≤ 1 year	14
34	22	21	21	because the interviewee feels that the info. she's providing is correct.	22	Diff. ≤ 1 year	22
35	33	34	34	spanish language issue	Not in a pair	Diff. ≤ 1 year	33
36	49	48	46	unable to correct	49	Diff. ≤ 1 year	49
37	22	21	22	?	Not in a pair	Diff. ≤ 1 year	22
38	22	23	23	verified age	22	Diff. ≤ 1 year	22
39	30	29	29	sr gave age wrong ra is actually 29 yrs of age	Not in a pair	Diff. ≤ 1 year	30
40	25	24	24	respondent was 24 yrs old on last bday now is 25	Not in a pair	Diff. ≤ 1 year	25
41	54	53	53	mistake	Not in a pair	Diff. ≤ 1 year	54
42	22	23	23	he confirms he is 23 yo	22	Diff. ≤ 1 year	22
43	46	47	46	A had use calculator to figure out his age. Told FI wrong age	46	Diff. ≤ 1 year	46
44	53	52	52	respondent is actually 53	Not in a pair	Diff. ≤ 1 year	53
45	45	44	44	would not provive day and month of birth	Not in a pair	Diff. ≤ 1 year	45

Table G.1 Explanations for Overrides to Consistency Checks against CURNTAGE (continued)

#	NEW AGE	Original Roster Age for Self	Screener Age	Verbatim Explanation from Field Interviewers[1]	Respondent's Age in Roster of Other Pair Member	Comment[2]	AGE = Final Roster Age
46	37	38	38	cannot get back to change the age from 37 to 38 ys	37	Diff. ≤ 1 year	37
47	33	32	33	Resp is 32 and husband is 33	Not in a pair	Diff. ≤ 1 year	33
48	42	43	42	mistake - didn't mean to his age is 42	42	Diff. ≤ 1 year	42
49	21	20	21	rostering person gave incorrect age for roommate	21	Diff. ≤ 1 year	21
50	14	13	14	Response from Interview Respondent	Not in a pair	Diff. ≤ 1 year	14
51	37	36	37	respondent turned 37 in october	37	Diff. ≤ 1 year	37
52	20	22	22	SR gave incorrect age	Not in a pair	***AGE was changed to equal roster age***	22
53	23	17	23	R gave incorrect age and dob. Error was discovered when ending interview.	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	23
54	36	9	30	This is a multi-family household	Not in a pair	Diff. > 1 year, FI does not support change leave as NEWAGE	36

[1] These entries came directly from the 2012 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and were not corrected.

[2] "Diff." refers to the difference between CURNTAGE and the age of the "self" in the household roster, the "roster age." Bolded and italicized entries indicate that the criteria for changing the age to that given in the household roster for "self" were met.

Table G.2 Explanations for Overrides to Internal Consistency Checks

#	Consistency Check	AGE	Roster Member's Age and Relationship to Respondent	Verbatim Explanation from Field Interviewers[1]	Comment
1	Gap of 25 or more years between brother and respondent	16	42-year-old brother	HH is 48, brother is 43	Overrule; unsure of relationship (85)
2	Gap of 25 or more years between sister and respondent	37	9-year-old sister	mother waited a while before next child	Overrule; probable child
3	Grandparent/Grandchild and respondent less than 30 years apart	17	3-year-old granddaughter	3	Overrule; unsure of relationship (85)
4	Multiple spouses or partners	27	42-year-old unmarried partner	male is girlfriend;female is girlfriend	Overrule; unsure of relationship (85)
5	Multiple spouses or partners	27	31-year-old unmarried partner	male is girlfriend;female is girlfriend	Overrule; unsure of relationship (85)
6	Respondent's daughter is less than 13 years younger than respondent	16	8-year-old daughter	rostered according to mother	Overrule; probable sibling
7	Respondent's daughter is less than 13 years younger than respondent	16	7-year-old daughter	rostered according to mother	Overrule; probable sibling
8	Respondent's daughter is less than 13 years younger than respondent	16	7-year-old daughter	rostered according to mother	Overrule; probable sibling
9	Respondent's daughter is less than 13 years younger than respondent	12	9-year-old daughter	9	Overrule; unsure of relationship (85)
10	Respondent's daughter is less than 13 years younger than respondent	16	6-year-old daughter	Child is 6yo	Overrule; unsure of relationship (85)
11	Respondent's daughter is older than respondent	17	20-year-old daughter	17	Overrule; probable sibling
12	Respondent's granddaughter is older than respondent	18	73-year-old granddaughter	oldest person 73-female-relationship granddaughter	Overrule; probable grandparent

Table G.2 Explanations for Overrides to Internal Consistency Checks (continued)

#	Consistency Check	AGE	Roster Member's Age and Relationship to Respondent	Verbatim Explanation from Field Interviewers[1]	Comment
13	Respondent's granddaughter is older than respondent	16	63-year-old granddaughter	63 YR ILD GRANDMOTHER TO RESPONENT	Overrule; probable grandparent
14	Respondent's mother is less than 13 years older than respondent	13	23-year-old mother	r says it is his mother	Overrule; unsure of relationship (85)
15	Respondent's mother is less than 13 years older than respondent	21	27-year-old mother	IR stated her mother's age is 28. IR is not certain about mother's age, but said to the best of her knowledge, that is her mother's age.	Overrule; unsure of relationship (85)
16	Respondent's son is less than 13 years younger than respondent	16	4-year-old son	Child is 4y/o second marriage	Overrule; unsure of relationship (85)
17	Respondent's son is older than respondent	13	14-year-old son	verified by mother	Overrule; probable sibling
18	Respondent's son is older than respondent	13	7-year-old son	verified by mother	Overrule; probable sibling
19	Respondent's son is older than respondent	13	5-year-old son	verified by mother	Overrule; probable sibling
20	Respondent's son-in-law is 16 or younger	26	5-year-old son-in-law	Per R, the 5y/o is his unmarried partners son.	Overrule; probable other nonrelative

[1] These entries came directly from the 2012 NSDUH field interviewers. Any typographical errors or misspellings were transcribed directly and were not corrected.

Appendix H: Rules for Determining Pair Relationships

Appendix H: Rules for Determining Pair Relationships

H.1 Rules for Determining Matching Pairs, in Priority Order

The following rules are used to determine the member of a respondent's household roster that corresponds to the other pair member. In these rules, an "age match" occurs if the questionnaire age of one pair member matches a roster age in the other pair member's roster, and a "gender match" occurs if the questionnaire gender of one of the pair members matches a roster gender in the other pair member's roster. In the table below, if the rules for Pair Member A and Pair Member B in a single row differ, then the count for that row includes the rules as listed, and the rules with Pair Member A and Pair Member B reversed. If the age and/or gender are off when finding these matches, the age and/or gender are defined by the questionnaire age and gender of the selected pair member when determining the pair domain. The rules, called match types, are listed in priority order in Table H.1, along with the number of pairs to which each rule was applied. Since the 2001 survey, it was technically impossible to identify more than one roster member as the "other pair member selected," resulting in either 0 or 1 MBRSEL for each responding pair. Rules involving situations where more than one MBRSEL existed are therefore not included in this table.

Table H.1 Rules for Determining Matching Pairs, in Priority Order

Match Type	Rule		
	Pair Member A	Pair Member B	Count
0.00	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, exactly one MBRSEL in right place	16,147
1.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	2,084
2.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, exactly one MBRSEL in right place	129
3.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	277
3.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	37
3.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within two, gender matches exactly, exactly one MBRSEL in right place	2

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
4.01	Age and gender match exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	280
4.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	21
4.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age and gender match exactly, MBRSEL missing for all roster members	6
4.05	Age and gender match exactly, MBRSEL missing for all roster members	Age and gender match exactly, MBRSEL missing for all roster members	17
4.06	Age and gender match exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	1
4.08	Age within one, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.09	Age within two, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.10	Age and gender match exactly, MBRSEL missing for all roster members	Two matches for age and gender, MBRSEL missing for all roster members, relationship is "Other Relative"	0
4.11	Age and gender match exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	1
4.13	Age within one, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0
4.14	Age within two, gender matches exactly, exactly one MBRSEL in right place	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0
4.15	Age and gender match exactly, MBRSEL missing for all roster members	Two matches for age and gender, MBRSEL missing for all roster members, roommate type relationship	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
5.01	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	23
5.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	4
5.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	0
5.05	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, exactly one MBRSEL in right place	1
5.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, exactly one MBRSEL in right place	0
8.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	41
8.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	6
8.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	0
8.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	2
8.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender matches exactly, MBRSEL missing for all roster members	0
8.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within one, gender matches exactly, MBRSEL missing for all roster members	0
10.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	164

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
10.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	22
10.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	7
10.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	6
10.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	1
10.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	2
10.11	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	Age within 10, gender matches, exactly one MBRSEL in right place, excludes cases where MBRSEL could have been applied to one of closer age	1
11.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	6
11.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	1

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
11.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
11.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	1
11.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
11.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within 10, gender matches, MBRSEL missing for all roster members, excludes cases where one of closer age could have been selected	0
12.01	Age and gender match exactly, exactly one MBRSEL in right place	Everything missing	17
12.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Everything missing	2
12.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Everything missing	0
12.05	Age and gender match exactly, MBRSEL missing for all roster members	Everything missing	0
12.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Everything missing	0
12.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Everything missing	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
13.01	Age and gender match exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	31
13.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	3
13.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	1
13.05	Age and gender match exactly, MBRSEL missing for all roster members	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	1
13.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	0
13.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Gender and reported household sizes match exactly, age missing, MBRSEL missing for all roster members	0
15.01	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	2
15.03	Age within one, gender matches exactly, exactly one MBRSEL in right place		0
15.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
15.05	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
15.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
15.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Multiple matches on age, gender, and relationship code; MBRSEL missing for all roster members; does not matter which match is picked	0
18.01	Age and gender match exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	1
18.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.05	Age and gender match exactly, MBRSEL missing for all roster members	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
18.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Multiple matches on gender, age within one, MBRSEL missing for all roster members, pick one where relationship code matches	0
19.01	Age and gender match exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	4
19.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
19.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0
19.05	Age and gender match exactly, MBRSEL missing for all roster members	Age within one, gender off, one MBRSEL, only two in household	0
19.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age within one, gender off, one MBRSEL, only two in household	0
19.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age within one, gender off, one MBRSEL, only two in household	0
19.11	Age within one, gender off, one MBRSEL, only two in household	Age within one, gender off, one MBRSEL, only two in household	0
20.00	No match, but no relationship codes are missing, and none involve domains of interest	No match, but no relationship codes are missing, and none involve domains of interest	24
21.01	Age and gender match exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.05	Age and gender match exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.06	Age matches exactly, gender off, exactly one MBRSEL in right place	Age matches exactly, gender off, MBRSEL missing for all roster members	0
21.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	Age matches exactly, gender off, MBRSEL missing for all roster members	0
22.01	Age and gender match exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	53

Table H.1 Rules for Determining Matching Pairs, in Priority Order (continued)

Match Type	Rule		
	Pair Member A	Pair Member B	Count
22.03	Age within one, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	8
22.04	Age within two, gender matches exactly, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	1
22.05	Age and gender match exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	2
22.06	Age matches exactly, gender off, exactly one MBRSEL in right place	No match at all (often paired respondent is missing from roster)	0
22.09	Age within one, gender matches exactly, MBRSEL missing for all roster members	No match at all (often paired respondent is missing from roster)	0
25.00	No match at all	No match at all	19

H.2 Rules for Identifying Pair Relationships among Pairs

Table H.2 summarizes the rules used to identify the pair relationships, using the relationship codes and questionnaire ages of the two pair members. Because the child (12 to 17)-parent and child (12 to 20)-parent relationships can be derived from relationships created using 12- to 14-year-olds, 15- to 17-year-olds, and 18- to 20-year-olds, these latter relationships are the ones referenced in the rules. The variable PAIRREL, which is the next to last column of the table, identifies the pair relationship as defined by Table 11.3 in the main body of this report. As with the rules for identifying which members of the roster belong to the pair, these rules—also called priority conditions because of their hierarchical nature—are shown in priority order. In the headers, the moniker "A" refers to pair member A, and "B" refers to pair member B. The relationship between A and B is described in the columns "A-B Relationship," from the perspective of pair member A ("B to A, according to A") and the perspective of pair member B ("A to B, according to B"). Any constraints on the pair members (other than FIPE3) are provided in the columns "Constraint on A" and "Constraint on B." These constraints include age constraints, where a range of ages (e.g., 12 to 17) indicates that the value of the questionnaire edited age (AGE) is between the numbers shown. Also in this column, "child" and "children" are defined as (a) roster member(s) with nonmissing ages less than 18. The question FIPE3 asks if the respondent is the parent of a selected 12- to 17-year-old. The responses provided in the table are either "yes" or "no." The column for RELMATCH indicates the quality of the match between pair members, as defined in Table 11.6 in the main body of this report. In the table, blank cells mean that no restrictions were placed on that variable to determine the pair relationship.

Table H.2 Rules for Identifying Pair Relationships among Pairs

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
1	Parent	Child	12-14				1	1
	Child	Parent		12-14				
2	Parent	Child	15-17				2	1
	Child	Parent		15-17				
3	Parent	Child	18-20				3	1
	Child	Parent		18-20				
4	Parent	Child	21+				4	1
	Child	Parent		21+				
5	Sibling	Sibling	12-14	15-17			5	1
	Sibling	Sibling	15-17	12-14				
6	Sibling	Sibling	12-17	18-25			6	1
	Sibling	Sibling	18-25	12-17				
7	Sibling	Sibling	No constraints, after considering #5 & #6				7	1
8	Spouse/partner	Spouse/partner	≥ 1 child	≥ 1 child			8	1
9	Spouse/partner	Spouse/partner	0 children, no bad data	0 children, no bad data			9	1
10	Spouse/partner	Spouse/partner	≥ 1 child	0 children, some bad data			8	1.5
	Spouse/partner	Spouse/partner	0 children, some bad data	≥ 1 child				
11	Spouse/partner	Roommate/nonrelative	≥ 1 child both sides, equal number each side				8	3
	Roommate/nonrelative	Spouse/partner	≥ 1 child both sides, equal number each side					

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
12	Partner	Partner	≥ 1 child	0 children, but other's children in household			8	3
	Partner	Partner	0 children, but other's children in household	≥ 1 child				
13	Spouse/partner	Spouse/partner	No constraints, after considering #8-12				10	1
14	Grandchild	Grandparent					11	1
	Grandparent	Grandchild						
15	Parent-in-law	Child-in-law					12	1
	Child-in-law	Parent-in-law						
	Other relative	Other relative						
	Roommate/boarder/ nonrelative	Roommate/boarder/ nonrelative						
16	Roommate/boarder/ other relative/ nonrelative/in-laws	Roommate/boarder/ other relative/ nonrelative/in-laws					13	1
17	Parent	Missing	12-14	B less than 10 yrs. older th. A		No	14	0
	Missing	Child						
18	Parent	Missing	12-14				1	2
	Missing	Child						
19	Child	Missing	A less than 10 yrs. older th. B.	12-14	No		14	0
	Missing	Parent						
20	Child	Missing		12-14			1	2
	Missing	Parent						
21	Parent	Missing	15-17	B less than 10 yrs. older th. A		No	14	0
	Missing	Child						
22	Parent	Missing	15-17				2	2
	Missing	Child						

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
23	Child	Missing	A less than 10 yrs. older th. B	15-17	No		14	0
	Missing	Parent						
24	Child	Missing		15-17			2	2
	Missing	Parent						
25	Parent	Missing	18-20				3	2
	Missing	Parent		18-20				
26	Child	Missing		18-20			3	2
	Missing	Child	18-20					
27	Parent	Missing	21+				4	2
	Missing	Parent		21+				
28	Child	Missing		21+			4	2
	Missing	Child	21+					
29	Sibling	Missing	12-14	15-17			5	2
			15-17	12-14				
	Missing	Sibling	12-14	15-17				
			15-17	12-14				
30	Sibling	Missing	12-17	18-25			6	2
			18-25	12-17				
	Missing	Sibling	12-17	18-25				
			18-25	12-17				
31	Sibling	Missing	No constraints, after considering #24, #25				7	2
	Missing	Sibling	No constraints, after considering #24, #25					
32	Spouse/partner	Missing	≥ 1 child	No spouse in roster			8	2
	Missing	Spouse/partner	No spouse in roster	≥ 1 child				

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
33	Spouse/partner	Missing	0 children, no bad data	No spouse in roster			9	2
	Missing	Spouse/partner	No spouse in roster	0 children, no bad data				
34	Spouse/partner	Missing	After #27, #28, no constraints	No spouse in roster			10	2
	Missing	Spouse/partner	No spouse in roster	After #27, #28, no constraints				
35	Grandchild	Missing	A at least 20 years older than B				11	2
	Missing	Grandparent						
	Grandparent	Missing	B at least 20 years older than A					
	Missing	Grandchild						
36	Roommate/boarder/ other relative/ nonrelative/in-laws	Missing			No		12	2
	Missing	Roommate/boarder/ other relative/ nonrelative/in-laws			No			
37	Roommate/boarder/ other relative/ nonrelative/in-laws	Missing					13	2
	Missing	Roommate/boarder/ other relative/ nonrelative/in-laws						
38	Nonmissing	Child	12-14			Yes	1	3
39	Nonmissing	Parent		12-14	Yes		1	3
40	Child	Nonmissing		12-14	Yes		1	3
41	Parent	Nonmissing	12-14			Yes	1	3
42	Nonmissing	Child	15-17			Yes	2	3

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
43	Nonmissing	Parent		15-17	Yes		2	3
44	Child	Nonmissing		15-17	Yes		2	3
45	Parent	Nonmissing	15-17			Yes	2	3
46	Parent	Roommate/boarder/ other relative/ nonrelative	12-14			No	13	3
						Missing	15	4
	Roommate/boarder/ other relative/ nonrelative	Parent		12-14	No		13	3
					Missing		15	4
47	Parent	Roommate/boarder/ other relative/ nonrelative	15-17			No	13	3
						Missing	16	4
	Roommate/boarder/ other relative/ nonrelative	Parent		15-17	No		13	3
					Missing		16	4
48	Parent	Roommate/boarder/ other relative/ nonrelative	18-20				17	4
	Roommate/boarder/ other relative/ nonrelative	Parent		18-20			17	4
49	Parent	Roommate/boarder/ other relative/ nonrelative	21+				18	4
	Roommate/boarder/ other relative/ nonrelative	Parent		21+			18	4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
50	Nonmissing, not sibling	Child	12-14	21-75		No	13	3
			12-14, exactly one parent	21-75, exactly one spouse		Missing	1	3
			12-14, 0 or 2 parents, or B has 0 or 2 spouse	21-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	15	4
	Child	Nonmissing, not sibling	21-75	12-14	No		13	3
			21-75, exactly one spouse	12-14, exactly one parent	Missing		1	3
			21-75, 0 or 2 spouses, or A has 0 or 2 parents	12-14, 0 or 2 parents, or B has 0 or 2 spouse	Missing		15	4
51	Nonmissing, not sibling	Child	15-17	24-75		No	13	3
			15-17, exactly one parent	24-75, exactly one spouse		Missing	2	3
			15-17, 0 or 2 parents, or B has 0 or 2 spouse	24-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	16	4
	Child	Nonmissing, not sibling	24-75	15-17	No		13	3
			24-75, exactly one spouse	15-17, exactly one parent	Missing		2	3
			24-75, 0 or 2 spouses, or A has 0 or 2 parents	15-17, 0 or 2 parents, or B has 0 or 2 spouse	Missing		16	4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
52	Nonmissing, not sibling	Child	18-20, exactly one parent	27-75, exactly one spouse		Missing	3	3
			18-20, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	17	4
	Child	Nonmissing, not sibling	27-75, exactly one spouse	18-20, exactly one parent	Missing		3	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	18-20, 0 or 2 parents, or B has 0 or 2 spouse	Missing		17	3
53	Nonmissing, not sibling	Child	21+, exactly one parent	27-75, exactly one spouse		Missing	4	4
			21+, 0 or 2 parents, or B has 0 or 2 spouse	27-75, 0 or 2 spouses, or A has 0 or 2 parents		Missing	18	3
	Child	Nonmissing, not sibling	27-75, exactly one spouse	21+, exactly one parent	Missing		4	3
			27-75, 0 or 2 spouses, or A has 0 or 2 parents	21+, 0 or 2 parents, or B has 0 or 2 spouse	Missing		18	4
54	Spouse	Sibling	One is 12-14, other is 15-17, both sides have parents or spouses				5	3
	Sibling	Spouse						
55	Spouse	Sibling	One is 12-17, other is 18-25, both sides have parents or spouses				6	3
	Sibling	Spouse						

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship: B to A, according to A	A-B Relationship: A to B, according to B	Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
56	Spouse	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, both sides have parents or spouses				7	3
	Sibling	Spouse						
57	Other relative	Sibling	Both sides have 2 parents, ages of oldest parents on either side differ by > 5 years, age of youngest parents on either side differ by > 5 years				13	3
	Sibling	Other relative						
58	Nonmissing, not child	Sibling	15-17	12-14			19	4
	Sibling	Nonmissing, not child	12-14	15-17				
59	Nonmissing, not parent	Sibling	12-14	15-17			19	4
	Sibling	Nonmissing, not parent	15-17	12-14				
60	Nonmissing, not child	Sibling	18-25	12-17			20	4
	Sibling	Nonmissing, not child	12-17	18-25				
61	Nonmissing, not parent	Sibling	12-17	18-25			20	4
	Sibling	Nonmissing, not parent	18-25	12-17				
62	Nonmissing, not child	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, A older than B				21 21	4
	Sibling	Nonmissing, not child	Ages neither 12-14/15-17 nor 12-17/18-25, B older than A					4
63	Nonmissing, not parent	Sibling	Ages neither 12-14/15-17 nor 12-17/18-25, B older than A				21 21	4
	Sibling	Nonmissing, not parent	Ages neither 12-14/15-17 nor 12-17/18-25, A older than B					4

Table H.2 Rules for Identifying Pair Relationships among Pairs (continued)

Priority Condition	A-B Relationship		Constraint on A	Constraint on B	FIPE3 (A)	FIPE3 (B)	PAIRREL	REL-MATCH
	B to A, according to A	A to B, according to B						
64	Sibling	Roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	At least one is between 18 and 20				13	3
	Roommate, in-law, grandparent, grandchild, boarder, other relative, nonrelative	Sibling	At least one is between 18 and 20					
65	Sibling	Unusual in-law code	12-20	26 or older			13	3
	Unusual in-law code	Sibling	26 or older	12-20				
66	Spouse/partner	Not a child, parent, or sibling	≥ 1 child aged < 18	No spouse			22	4
	Not a child, parent, or sibling	Spouse/partner	no spouse	≥ 1 child aged < 18				
67	Spouse/partner	Not a child, parent, or sibling	15 or older, 0 children, no bad data	15 or older, no spouse			23	4
	Not a child, parent, or sibling	Spouse/partner	15 or older, no spouse	15 or older, 0 children, no bad data				
68	Grandparent, grandchild	Not grandparent, not grandchild					25	4
	Not grandparent, not grandchild	Grandparent, grandchild						
69	Any codes	Any codes	No constraints	No constraints			14	0

Appendix I: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

Appendix I: Conditions Used for Reconciling Differing Multiplicity Counts between Pair Members

I.1 Introduction

In order to determine multiplicity counts, counts were obtained from each pair member. The count from the pair member who was the focus member of the domain was considered the direct count, and the count from the other pair member was considered the indirect count. Typically, these counts were in agreement, and the determination of the final multiplicity count was straightforward, provided neither roster had bad data codes. The strategy also was usually clear if one pair member had bad data in the household roster, or had a 0 count when the pair relationship precluded a value of 0. The count from the pair member with good, nonzero data was usually preferred in those cases. If the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. The rules that were used to reconcile these disagreeing counts are outlined in this appendix.

Note that the reconciliation of differing counts was necessary for parent-child and sibling-sibling pairs but was not necessary for spouse-spouse pairs, since the multiplicity count for spouse-spouse pairs was always 1. As noted in Section 11.3.1, it was technically possible for a respondent to have multiple spouses, but these situations were not accounted for.

I.2 Parent-Child Counts

For parent-child counts, the screener and the FIPE3 variable were used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. The multiplicity counts in this domain reflected the selected child's parents and in most cases had values of 1 or 2. If neither side had bad relationship codes, and if the direct count exceeded the indirect count, the following rules applied:

1. The direct count might have exceeded the indirect count because one parent had left or entered the household between interviews. In this case, the ages in the rosters were matched to the screener roster to determine which count to believe. This was done in two ways. First, the total number of roster members between ages 30 and 39, 35 and 44, and 40 and 49 were compared between pair members and the screener. The pair member with age range counts closest to the screener was the one whose parent-child count was chosen for the final count. If neither side had age range counts equal to the screener, then the pair member with a parent-child count equal to the total number of screener roster members between ages 26 and 64 was chosen as the final count.

2. The direct count might have exceeded the indirect count because the selected parent did not consider the other "parent" a spouse or live-in partner. If the pair relationship was not imputed, the indirect count was selected. However, if the pair relationship was imputed and the older pair member called the younger pair member a child, then the older pair member considered the child's "true" parent as not a spouse or live-in partner, even though he or she claimed the "true" parent's children. In this case, the direct count was used (the child's adjusted count).

If the direct count was exceeded by the indirect count, then the child listed only one parent, and the parent listed a spouse (a "stepparent") or live-in partner in the household roster. The following rules applied:

1. The indirect count might have exceeded the direct count because the selected child did not accept a stepparent or live-in partner as his or her parent. If this stepparent or live-in partner was the other respondent selected, we determined that this was a child-parent pair based on the response of the "parent" to the FIPE3 question. If the FIPE3 question was answered "yes," and the RELMATCH variable had a value of 3, then the indirect count was selected as the multiplicity count. If the FIPE3 question was answered "no," the pair was not considered a child-parent pair and was not considered for these counts. Finally, if the FIPE3 question was not answered, the respondent was considered a "parent" if he or she was a stepparent. If the respondent was a live-in partner, the determination of the pair relationship was left to imputation. The multiplicity count was set to the indirect count to account for the possibility that the pair relationship would be imputed as parent-child.

2. Suppose the selected child did not accept a stepparent or live-in partner as his or her parent (as above), but the other respondent selected was the "true" or "original" parent. In this case, the stepparent or live-in partner was identified only in the "original" parent's roster, so there was no way to determine how the stepparent or live-in partner would have answered the FIPE3 question. A stepparent was considered a "parent" even if the child did not view him or her this way so that the indirect count was used. The case of live-in partners was less clear, and these cases were left to imputation.

3. If age range counts between the two pair members and the screener matched across a variety of age ranges (30 to 39, 40 to 49, and 50 to 59), but the child's roster had a bad relationship code among roster members of potential parent age (15 or older), or the child's roster had a value of MBRSEL that did not match what was finally determined to be the child's parent, then the multiplicity count for the parent—the indirect count—was selected as the final count.

Parent-child pairs, parent focus. The multiplicity counts in this domain reflected the selected parent's children and were limited to have values of at least 1. If neither side had bad relationship codes, the following rules applied:

1. If the count of children in the household within the relevant age ranges differed between the pair members, but one side had a count of children equal to the same

count from the screener roster, then the multiplicity count that corresponded to the pair member with the same count of children as the screener was used.

2. If the count of children in the household within the relevant age ranges differed between the pair members, and both sides had a multiplicity count that exceeded the count of all children from the screener roster, then the number of children in the screener roster was used as the multiplicity count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

3. The direct count and indirect count might differ because either the child listed a sibling that the parent considered "another relative" or the parent listed a child that the child considered "another relative." In either case, the parent was the one to answer the FIPE3 question. Because of this, the multiplicity count from the parent's perspective was selected as the final count, provided that the counts of children in the household within the relevant age ranges for each pair member were equal.

4. After considering the above situations, the multiplicity counts might have still differed without a resolution of which count should have been chosen as the final count. This occurred because the counts of children in the household differed between pair members, each of which differed from the screener count. Moreover, multiplicity counts did not exceed the screener age range count. In this instance, if one of the multiplicity counts equalled the screener age range count, then this multiplicity was selected as the final count. However, if this was not the case, then upper and lower bounds were created and the final multiplicity was left to imputation.

Because of the hierarchical nature of these counts, parent-child counts for 12- to 17-year-old and 12- to 20-year-old children could sometimes be derived if the 12- to 14-year-old parent-child count was already determined for both child focus and parent focus counts. In particular, if one pair member's count for 12- to 17-year-old children or 12- to 20-year-old children equalled or exceeded the final parent-child count for 12- to 14-year-old children and the other did not, then the pair member's count that equalled or exceeded the 12- to 14-year-old count was chosen as the final count.

I.3 Sibling-Sibling Counts

Although there were two types of sibling-sibling pairs under consideration, each associated with two domains, the same rules could be applied to all four domains. When the older sibling was the focus, the multiplicity count was a count of the number of siblings within the younger age group (12 to 14 or 12 to 17). Conversely, the multiplicity count was the number of siblings in the older age group (15 to 17 or 18 to 25) when the younger sibling was the focus. Deciding how to assign a final multiplicity count often involved looking at a count of household members within the age range of the siblings being counted. For example, if the older sibling was the focus and the age ranges were 12 to 14 and 15 to 17, the number of household members aged 12 to 14 were counted. The following general rules applied if the multiplicity counts for each pair member disagreed:

1. The counts disagreed if a household member left or entered the household between interviews. As before, the roster that was closest to the screener was used to determine the count. In particular, depending upon the domain, the count of household members within the age range of the siblings being counted was compared between each pair member and the screener. The multiplicity count from the pair member with the count closest to the screener was used, provided that the member had no bad relationship codes within the relevant age range.

2. If the counts of household members within the age range of the siblings being counted differed between pair members and those counts were both exceeded by the screener count, then the multiplicity associated with the pair member with the age range count closest to the screener was chosen, provided that the member had no bad relationship codes within the relevant age range.

3. In some cases, the counts of household members within the age range of the siblings being counted were the same for the two pair members, but the multiplicity counts disagreed.

    a. If one pair member had bad relationship codes and the other did not, the disagreement could have been due to the bad relationship codes. If the sums of the multiplicity count and the number of bad relationship codes were equal across pair members, then the final count was set to equal the multiplicity of the pair member who did not have bad relationship codes.

    b. If one pair member identified the other as "sibling" but the other pair member did not reciprocate, then imputation was required to establish whether the relationship was sibling-sibling. The count associated with the pair member who indicated that the other pair member was a sibling should have been chosen as the final count. In effect, this was done by taking the maximum of the two pair members' counts.

4. If the counts of household members within the age range of the siblings being counted disagreed and both exceeded the screener count of household members within the relevant age range, then the multiplicity count was set to the screener count. If the screener roster had missing exact ages, then the minimum multiplicity count from the two pair members' rosters was used as the final count.

5. If the differing multiplicity counts could not be reconciled with the above rules, upper and lower bounds for the true multiplicity were determined using the two multiplicity counts, as well as the counts of children within relevant age ranges in both pair member's rosters and the screener roster. In rare cases, the values for these bounds were equal. These cases were investigated, and if the reasons were legitimate, then the final multiplicity count was set to this value. Otherwise, the final multiplicity was left to imputation.

Appendix J: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

Appendix J: Conditions Used for Reconciling Differing Household-Level Person Counts between Pair Members

J.1 Introduction

Household-level person counts for a particular domain were obtainable using the multiplicity counts if the pair belonged to a pair relationship that fit into that domain, provided only one family unit was in the household. No reconciliation between pair members was necessary in that case, since the reconciliation had already been done with the multiplicity counts. Other counts were obtained from single respondents for whom no reconciliation was necessary. This appendix discusses the conditions used to reconcile differing household-level person counts when the pair belonged to a pair relationship that corresponded to different pair domains than the one being counted. Typically, the counts between the two pair members were in agreement, and the determination of the final household-level count did not involve a reconciliation of counts, though assigning a final count meant ensuring that pair relationships were not hidden due to the relationships of the two pair members to other household members.[1] A similar situation occurred if one pair member had bad data in the household roster. The count from the pair member with good data was usually preferred in those cases, provided pair relationships of interest were not hidden. If bad data existed in either household roster, but the bad data was limited to bad relationship codes, then the member with good data was selected only if substituting the appropriate relationship codes for the bad data codes would have given a total that was equal to the count from the pair member with good data. There were instances where bad data codes existed in the roster, and this condition did not apply. There were other exceptions as well. Finally, there were instances where neither pair member had bad data in their rosters, yet their counts still disagreed. In this appendix, the rules that were used to assign a final count, as well as to reconcile disagreeing counts, are outlined. For each pair domain, a set of general rules are given, each with specific conditions required for the general rule to be implemented. Within each general condition, if at least one of the specific conditions was not satisfied, upper and lower bounds were determined and the final count was left to imputation.

J.2 Parent-Child Counts

For parent-child counts where the pairs were not parent-child pairs of interest (e.g., sibling-sibling pairs, parent-child pairs where the child was 21 or older, etc.), the screener was used to help reconcile disagreeing counts. The rules follow below, separated by the member of focus:

Parent-child pairs, child focus. For the child-focus counts, the count is of the number of children of a parent in the household. The following general rules applied:

1. In most cases, the counts of children in the relevant age range with parent(s) in the household (abbreviated below as children with parent(s) in the household) for the two

[1] If a roster pointed to a household size of one, this was considered "bad data" since both pair members in the household were survey respondents.

sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:

- Either no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or the counts of children with parent(s) in the household were equal to the screener age counts, or a side with good data indicated siblings within the relevant age range living together in a household without parents;
- No situations where parents were not identified in the household, but some in the household had bad relationship codes and were old enough to be parents;
- No counts of one child in the relevant child-age range when both members of the pair were in that range and the children were siblings;
- No pairs where the ages of the identified parents did not match, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household;[2] and
- The household size was greater than 1 and was nonmissing on both sides.

2. The counts of children with parent(s) in the household might have agreed even though the above conditions were not met. The final count of children with parent(s) in the household could still have been set to one of the sides, if any one of the following was true:

   - If the number of children within the relevant age ranges matched across both rosters and the screener and (at least) one side had all good age and relationship codes, provided the equal counts did not refer to different children;[3]
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and (at least) one side had all good age and relationship codes;
   - If both sides had a count of zero children with parent(s) in the household, both had a roster, and the number of respondents who were old enough to be parents in the household was zero according to the screener; or
   - If the counts of children with parent(s) in the household that agreed with each other equalled or exceeded the count of the number of children from the screener within the relevant age ranges.

3. The counts of children with parent(s) in the household might have agreed with a value of 1. If both pair members were children within the relevant age range, and both indicated they had parents even though the children were siblings, then they were not included in each other's rosters, but they were obviously in the screener roster, so the final count of children with parent(s) in the household was set to 2.

---

[2] Codes that indicate "other relative" or a nonrelative are 7 (roommate), 8 (child-in-law), 10 (parent-in-law), 12 (boarder), 13 (other relative), and 14 (other nonrelative).

[3] This was determined by excluding situations where the ages of the identified parents did not match, the pair members were not siblings, and both sides had relationship codes signifying "other relative" or a nonrelative, indicating more than one family unit in the household.

4. If one pair member did not have a valid roster but the other member did, the final count of children with parent(s) in the household was set to the other pair member's count under the following conditions:
    - No counts of one child with parent(s) in the household when both members of the pair were children in the relevant age range and the children were siblings, and
    - Either:
        - There were no bad relationship codes within the relevant child-age ranges and the respondent identified parents or children in the household,
        - There were no children within the relevant age range, or
        - No parents were identified in the household and nobody in the roster older than the respondent had a bad relationship code.
5. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count of children with parent(s) in the household, if that count was 0, under any of the following conditions. Either:
    - The other roster was valid, did not have any bad ages, and had no ages in the relevant age range;
    - The other roster also was bad but the screener roster was valid and did not have any ages in the relevant age range; or
    - The respondent identified both grandchildren and grandparents in the roster where the "grandchild" relationship code(s) were incorrectly entered into the respondent's household roster. The "grandchildren" that these relationship codes were referring to were not the respondent's grandchildren, but, rather, they were the respondent's grandparent's grandchildren.[4]
6. When two different family units were in the household, the determination of the final count of children with parent(s) in the household had to be treated separately. This could have included the multigenerational families referred to earlier and the two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts of children with parent(s) in the household (one count might be 0) was used as the final count, provided the following conditions were satisfied on both sides:
    - There were no bad ages or relationship codes within the relevant age ranges;
    - Both had counts of children with parent(s) in the household pointing to two or fewer parents, meaning that the two family units were not identifiable on one side;

[4] This condition has not manifested itself since the 2001 survey. With the addition of a new consistency check added since the 2001 survey to address grandparent/grandchild code inconsistencies, this condition could be observed only if a respondent overrode this consistency check, which has not happened.

- The number of identified parents was not equal to the total number of household members older than 25 in the household on either side, meaning that parents could correspond to roster members identified by other relationship codes;
- The number of identified children was not equal to the total number within the relevant age range in the household on either side, meaning that children with parents could correspond to roster members identified by other relationship codes; and
- There were not three generations in the household with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

If the pair was a parent-child pair, the final count was determined using imputation.

7. Two family units might be in the household but the conditions given in item #6 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of children with parent(s) in the household were determined for each side. The maximum of the two counts was used as the final count if the household members in the roster older than 25 (of parental age) were either both equal to the number of household members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of children with parent(s) in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.
8. If one pair member did not have a valid roster and the pair member with a valid roster was within the valid age range and was a sibling to the other pair member, but the count of children with parent(s) in the household from his roster was only 1, then the final count was set to 2.
9. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, it was necessary to decide who to believe. This occurred often because one of the respondents was a relative outside the nuclear family unit—like a cousin or aunt/uncle—whose own parents did not live in the household, or the respondent was a boarder.[5] Selecting either the zero count or nonzero count in this instance required that the following conditions were met:
   - The respondent with a zero count of children with parent(s) in the household did not identify parents in the roster or he or she identified parents but was older than 20 and had no bad relationship codes within the relevant age ranges, and

[5] Even if there was disagreement between the respondents about whether a boarder or other family member was in fact a sibling, parent, or child, this would had been resolved at the pair relationship stage where we would had determined whether this was in a domain of interest.

- Either the respondent with a nonzero count of children with parent(s) in the household had siblings or children within the relevant age range, or the respondent himself or herself was within that age range (with a count of 1).

When one count of children with parent(s) in the household was zero and the other was nonzero, the nonzero count was used under the following conditions:

- The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges, and
- Either:
  - The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, and the count of children with parent(s) in the household did not exceed the screener count of children within the relevant age range;
  - The count of children in the household within the relevant age range for the nonzero count pair member matched that of the screener;
  - The count of children in the household within the relevant age range for the zero count pair member matched that of the screener because a child was (or children were) listed as 11 years old in the nonzero count pair member's roster, when he or she (they) should have been 12 (according to the zero count pair member's and the screener roster) so that the final count was the nonzero count with this child (these children) added;
  - The respondent with a zero count had no household members with a family-type relationship code and the reported household sizes of the two pair members were equal (indicating that it was unlikely that anyone had entered or left the household between interviews);
  - The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within the relevant age range in the household for the zero count was closer to the screener age count. Nevertheless, the nonzero count was equal to or less than the screener age count; or
  - The other conditions had not already established a nonzero count, but a count for a subset age group had already been established as nonzero. For example, if the count for 12- to 14-year-olds was nonzero, then the 12- to 17-year-old count had to be nonzero.

The zero count of children with parent(s) in the household was used if the zero-count respondent had no bad relationship codes at all, and either:

- The household age composition among the relevant age ranges for the zero count pair member more closely matched the screener, or
- The pair was a grandparent-grandchild pair with an adult child of the grandparent living in the household. The nonzero count resulted from an assumption that a respondent's adult child and grandchild within the relevant age range were a parent-child pair. If the grandchild identified the grandparent's child as "other

relative" and did not identify any parents, this indicated that the grandparent's adult child was an uncle/aunt of the grandchild, not a parent.

10. If the pair relationship was not parent-child nor was it sibling-sibling, but one side had nonzero counts of children with parent(s) in the household and the other did not, taking the side that was closest to the screener sometimes meant that the count of children with parent(s) from neither pair member was chosen. As with the previous item, a zero count and a nonzero count often occurred because one of the respondents was a relative outside the nuclear family unit—like a cousin or aunt/uncle—whose own parents did not live in the household, or the respondent was a boarder. If neither the zero count nor the nonzero count was chosen, the final count could still have been determined using either the screener count, the count of children within the relevant age range for the respondent with a zero count, or one less than the nonzero count. One of these was chosen, provided that the following conditions were met:

    - The respondent with a zero count of children with parent(s) in the household did not identify parents in the roster or he or she identified parents but was older than 20 and had no bad relationship codes within the relevant age ranges, and
    - Either the respondent with a nonzero count of children with parent(s) in the household had siblings or children within the relevant age range, or the respondent himself or herself was within that age range (with a count of 1).

    The screener count was chosen if either:

    - The respondent pair member with a nonzero count also did not have bad relationship codes within the relevant age ranges. The count of children within the relevant age range in the household for the nonzero count pair member matched that of the zero count pair member, and the nonzero count exceeded the screener count of children within the relevant age range.
    - The respondent with a nonzero count showed a parent-child relationship existed in the household, but the respondent with a zero count did not because he was not related to the other household members. However, the count of children within the relevant age range in the household for the respondent with the zero count was closer to the screener age count, and the nonzero count exceeded the screener count of children.

    In situations where a respondent with a zero count had a roster more closely resembling that of the screener, but the screener included a household member within the relevant age range who was not part of the immediate family, neither the nonzero count of children with parent(s) in the household nor the screener count of children within the relevant age range could be used—a different count had to be used. Two strategies were employed:

    - For the respondent with a nonzero count of children with parent(s) in the household, the nonzero count was the same as the count of children within the relevant age range in the household, but it exceeded the number-of-children count for the zero-count respondent. However, the count of children within the relevant

age range for the zero-count respondent, which was not zero, was closer to the screener age count than the nonzero-count respondent.

- If the count of children within the relevant age range for the zero-count respondent was the same as the nonzero count of children with parent(s) in the household, the number-of-children count for the zero-count respondent could not be used, since the nonzero count included a household member that was not in the appropriate age range at the time of screening. One less than the nonzero count of children with parent(s) in the household was therefore chosen as the final count.

11. Other situations with a zero and nonzero count did not necessarily mean that the relationship was something other than parent-child or sibling-sibling. This was usually due to one pair member having missing relationship codes for the roster member that would have been identified as a parent (i.e., relationship codes for roster members in a parental age range). If the count for the pair member with the entirely good roster was equal to the number within the relevant child age range for the pair member with bad relationship codes in the roster, the nonzero count was selected.

12. The two counts of children with parent(s) in the household might have disagreed where both were nonzero and both exceeded the screener count of children within the relevant age range. For the screener count to be chosen as the final household count of children with parent(s) in the household, the following conditions had to be met:

    - The pair member's household rosters had to have different numbers of children within the relevant age range,
    - The pair relationship could be neither parent-child nor sibling-sibling with a zero screener count of children within the relevant age range(s),
    - The total number within the screener roster (where the minimum age was 12 years) had to be at least two, and
    - The number of children in the screener roster within the relevant age range was valid and at least as large as the final count of children with parents in the household for the next smallest age range.

13. The two counts might have disagreed because one side had bad relationship codes within the relevant age range and the other did not. If the sum of the number of bad relationship codes with the smaller count equalled the larger count, the larger count was chosen.

14. The two counts might have disagreed because they disagreed on the ages of one or more household members, even though each respondent's count included all the children in their respective roster. If the roster for one respondent more closely matched the screener in terms of the distribution of ages within the roster, then that respondent's count was chosen.

15. The two counts might have disagreed because they disagreed on the ages of one or more household members and each respondent's count included all the children in their respective roster, but neither was closer to the screener count. If the screener count differed from each respondent's count by the same amount, was greater than 1 but less than the other, then the screener count was used as the final count.

16. If the pair relationship was parent-child and the parent-child counts were associated with the same age range, then the household-level person counts were obtained using the parent-focus multiplicity counts corresponding to the appropriate age range. However, this did not occur if the age range for the pair relationship differed from the age range for the parent-child counts. If the pair relationship was imputed to be parent-child or it was deemed parent-child even though the child did not consider the parent a "parent," but the parent answered the FIPE3 question, then the nonzero count was used as the final count.

17. If, after all the above tests were done to find the final count, the minimum possible and maximum possible counts—considering both questionnaire rosters and the screener roster—were the same, then the final count was set to that value.

18. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

Parent-child pairs, parent focus. For the parent-focus counts, the count is of the number of parents of at least one child in the household. The child-focus parent-child counts are processed first, so if the child-focus parent-child counts are 0, it necessarily means that the parent-focus counts will also be 0. Nonzero child-focus counts also point to nonzero parent-focus counts. After setting counts to 0 where necessary, the following general rules applied:

1. In most cases, the counts of parents with children in the household for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:

   - No situations where both pair members were children in the relevant age range but were in a spouse-spouse pair relationship and both identified the same roster member as a parent,
   - The household size was greater than 1 and nonmissing on both sides, and
   - Either:
     - No bad relationship codes for household members of an age to be parents,
     - The total count was 2 for two parents, or
     - The total count plus the number of grandparents equalled the total number of household members aged 26 or older, according to the screener roster.

   Note that it was not necessary to check for bad relationship codes in the child age ranges, since it was already known that the count had to be at least 1, and the number of children was not important for the parent counts.

2. The counts of parents with children in the household might have agreed even though the above conditions were not met. The final count could still have been set to one of the sides if it was a sibling-sibling pair, and the bad codes in the parental age range were on one side only. This would indicate that the side with bad codes were not missing parental codes.

3. If one pair member did not have a valid roster but the other member did, the final count of parents with children in the household was set to the other pair member's

count if there were no bad relationship codes and no roster members with bad age and bad gender values. Other circumstances called for setting the final count to 0, which would necessarily be the case if the child-focus counts were 0.

4. When two different family units were in the household, the determination of the final count of parents with children in the household had to be treated separately. This could have included multigenerational families or two siblings both with children in the relevant age range living in the household. The latter was more easily identified if it was not a parent-child pair (e.g., a cousin-cousin pair). The sum of the two counts (one count might be 0) was used under the following conditions:

   - There were no bad ages or relationship codes within the relevant age ranges,
   - Both pair members had counts pointing to 2 or fewer parents, meaning that the two family units were not identifiable on a side,
   - The number of identified parents was not equal to the total number of household members older than 25 on either side, meaning that parents could correspond to roster members identified by other relationship codes, and
   - There were not three generations in the household, with first and second generation parents both having children in the appropriate age range. This was already accounted for by the counts for one or both sides.

5. Two family units might be in the household but the conditions given in item #4 were not met. If there were no bad ages or relationship codes within the relevant age ranges (for both children and parents), the two families in the household might have been already accounted for when the counts of parents with children in the household were determined for each side. The maximum of the two counts was used as the final count if the household members older than 25 (of parental age) in the roster were either both equal to the number of members older than 25 in the screener roster or both different than the number of members older than 25 in the screener roster. However, if the number of household members older than 25 in the screener roster was equal to the number of members older than 25 in one of the pair member's rosters but not the other, then the count of parents with children in the household corresponding to the pair member with a roster matching the screener roster (among household members of potential parental age) was used as the final count of children with parent(s) in the household.

6. If the pair relationship was a spouse-spouse pair and one of the pair members was within the relevant age range and had a positive count, then the count for that pair member was taken as the final count, provided there were no bad relationship codes in that roster for roster members aged 18 or older.[6]

7. The two counts might have disagreed with one nonzero count and the other equal to zero. Due to the fact that the counts of parent(s) in the household with children were determined first and that the zero counts were handled separately, the final count of parents with children in the household determined at this stage of processing had to

[6] For this condition, either the count for the other pair member was 0 or the count for the pair members was equal.

be nonzero. Counts arising from two or more families in the household also were handled in previous code. Hence, the final count had to be one or two parents.[7] The nonzero count was chosen as the final count if one of the following conditions were met:

- The count was 1 and there were no bad ages with the relevant relationship codes and no bad relationship codes within the relevant age ranges, or
- The count was 2.

8. The two counts might have disagreed where the number of roster members aged 26 or older disagreed between the two pair members. In these situations, one count was 1, and the other count was 2. The final count corresponded to the pair member with the number of roster members aged 26 or older closest to the screener number of roster members aged 26 or older, under the following conditions:
   - The difference between the screener count of the number of household members aged 26 or older and the pair members' counts of this number of household members was not the same between the two pair members,
   - Neither pair member had bad ages in their rosters, and
   - Each pair member either had no bad relationship codes in his or her roster or had a nonzero count with no bad relationship codes among respondents aged 26 or older.
9. The two counts might have disagreed if the bad relationship codes referred to missing parental codes. If one side had no bad relationship codes, and the sum of the number of bad relationship codes and the count on the side with the bad codes was equal to the count on the side with no bad relationship codes, then the count from the side with no bad relationship codes was used as the final count.
10. The two counts might have disagreed where one count was 2 and the other was 3. Since households with two family units had already been considered, the maximum number of parents possible was two, so the final count was set to 2.
11. If the pair relationship was parent-child and the parent-child counts were associated with the same age range, then the household-level person counts were obtained using the child-focus multiplicity counts corresponding to the appropriate age range.
12. If, after all the above tests were done to find the final count, the minimum possible and maximum possible counts—considering both questionnaire rosters and the screener roster—were the same, then the final count was set to that value.
13. Remaining disagreeing counts were left to imputation, with appropriate bounds set on the imputed value.

[7] This precluded the extremely unlikely possibility that the pair member with a zero count masked a situation where three parents in a single family unit lived in the household (two biological parents and a stepparent).

J.3 Sibling-Sibling Counts

The logic for the sibling-sibling counts did not depend upon whether the lower age range was 12 to 14 or 12 to 17 or whether the upper age range was 15 to 17 or 18 to 25. It also did not depend upon which pair member was the focus, though for the household-level person counts, the older member focus counts were the only ones considered. Hence, the counts of interest are of roster members in the upper age range. As with the parent-child pairs, the multiplicity counts could be used if the pair relationship was a sibling-sibling pair of interest. However, the counts had to be determined for all other pairs. The rules follow below, separated by the member of focus:

1. In most cases, the counts for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:
   - The pair could not be a sibling-sibling pair, where both respondents were in the upper age range, and could not have a younger sibling in the lower age range, and the count was 1. (This refers to a sibling-sibling pair that would not constitute a domain of interest.)
   - No bad relationship codes in the lower range if the count was 0.
   - Either:
     - No bad relationship codes in the upper range, or
     - The count matched the screener age count.
   - The household size was greater than 1 and nonmissing on both sides.
2. The counts might have agreed even though the above conditions were not met. The count could still have been set to one of the sides if any one of the following conditions was true:
   - If the number of children matched across both rosters and the screener for both the upper and lower age ranges, or
   - If the count was 0 and one of the following two conditions was true:
     - Neither side had bad relationship codes or ages, or
     - The number of household members aged 26 or older in the screener roster was zero.
3. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under the following conditions:
   - No bad relationship codes within the lower age range when the count was 0.
   - Either:
     - There were no bad relationship codes within the upper age range,
     - The count was equal to the screener age count within the upper age range, or
     - The count was 0, and the count of household members in the lower age range was 0.

4. If one pair member did not have a valid roster but the other member did, and the above conditions were not met, it was still possible to use the other pair member's count under the following conditions:
    - The count was 0,
    - The number of children in either the lower or upper age ranges was 0 with no bad ages in the roster.
5. If neither pair member had a valid roster, it was occasionally still possible to assign a final count. If the number of children in the screener roster in either the lower or upper age ranges was zero and the screener roster was valid, then it was not possible for a sibling-sibling pair in the relevant age ranges to be selected and the final count to be set to 0.
6. When two different sets of siblings were in the household, the determination of the final count had to be treated separately. The two sets of siblings refer to siblings where both parents from one set differ from the parents of the other set. The sum of the two counts (one count might be 0) was used, provided the following conditions were satisfied for both pair members:
    - The sum of counts of the number of sibling-sibling pairs equalled or exceeded at least one of the counts of household members in the upper age range for the screener roster or either of the pair member's rosters.
    - There were no bad relationship codes within the upper age ranges.
    - There were no bad relationship codes within the lower age range, or the count was nonzero.
7. If the counts from the two pair members did not agree, the following rules were used to assign the appropriate count, provided no bad relationship codes were evident in either age range on either side. These conditions are hierarchical, in that subsequent conditions require that the previous condition was not met.
    - If the number within the upper age range was the same on both sides, but the number in the lower age range was not, then the side with the number in the lower age range equal to the number in the screener roster within the lower age range was chosen. (In all cases, one side had a zero count and the other did not. This captured situations where it was necessary to discern whether the zero count was due to no children in the lower age range on one side and whether the screener also had no children in that range.)
    - For one pair member, the number of children in either the lower age range or the upper age range did not agree with the number in the screener roster in that range. However, for the other pair member, the number within both age ranges agreed with the screener count. The count was set to the side that agreed with the screener.
    - For both pair members, the numbers within the lower age range were either both zero or both positive. The number within the upper age range did not agree

between pair members, but one pair member agreed with the screener. The final count was set to the count for that pair member.

- In the rosters for both pair members and the screener, the numbers within the upper age range for at least one of the three were nonzero but not necessarily equal. The numbers within the lower age range were not equal across any of the three rosters. The pair member with the number of children in the upper age range closest to the screener was selected.

8. If the counts from the two pair members did not agree, but one side had bad relationship codes within the upper age range and the other did not have bad relationship codes, and the sum of the count and the number of bad relationship codes on one side was equal to the count for the pair member with the good roster, then the count for the pair member with the good roster was selected.

9. If the counts from the two pair members did not agree, and the above conditions were not met, in many cases this was due to one of the pair members not being part of the immediate family unit, in which case his or her count was automatically 0. To identify these cases and assign the count to the other pair member, the following conditions had to be satisfied:

   - The pair relationship did not indicate an identifiable family-type relationship (e.g., sibling-sibling, parent-child, spouse-spouse, or grandparent-grandchild relationship).
   - Either:
     - One pair member did not have any relationship codes indicating parent, child, sibling, spouse, grandchild, or grandparent;
     - The other pair member had at least one relationship code indicating a relationship other than parent, child, sibling, spouse, grandchild, or grandparent;
     - For the pair member with family codes, either no bad relationship codes were within both the upper and lower age ranges or no bad relationship codes were within the upper age range, and the count was positive; or
     - There were no bad relationship codes within both the upper and lower age ranges for either pair member.

10. If one pair member had no bad relationship codes within both the upper and lower age ranges, but the other member had some bad codes, then the count associated with the pair member with no bad codes was selected if the count of immediate family members (parent, child, sibling, spouse, grandchild, or grandparent) was the same as the count of household members within both the lower and upper age ranges.

11. If one pair member had a zero count due to having no household members within the upper age range, but the number of household members within that age range was nonzero for both the screener and the other pair member (though not necessarily equal), and the count for the other pair member was equal to the number of household members within the upper age range for that pair member, then a nonzero count was selected. If the number of household members within that age range in the screener

roster was nonzero, then that number was chosen as the final count. Otherwise, the number of household members within the upper age range for the pair member with nonzero count was selected as the final count.

12. If the pair was a spouse-spouse pair, one count might have been zero while the other was nonzero because the spouse-spouse pair still lived with the parents of one pair member, and the pair member's younger siblings also lived in the household. In this case, the nonzero count was selected if the number of immediate family members (parent, child, sibling, spouse, grandchild, or grandparent) in the roster for the pair member with the zero count was less than his or her total household size.

13. In some cases, one pair member called the other pair member a parent or child, but the other pair member did not reciprocate. In the case of a child who did not reciprocate the parent's identification of him or her as a child, the child's count was always less than the parent's count. By the same token, in the case of a parent who did not reciprocate the child's identification of him or her as a parent, the parent's count was always less than the child's count. If the pair relationship was imputed to be "parent-child," then the pair member who did not acknowledge a parent-child relationship was overruled, and the maximum count of the two pair members was selected as final.

J.4 Spouse-Spouse Counts (with or without Children)

The multiplicity counts were not useful in the logic for the spouse-spouse household counts, since the spouse-spouse multiplicity counts were always 1.[8] If the household size was one, or the number of respondents aged 15 or older in the household was one or zero, then the final household person count was set to 0 since no spouse-spouse pairs could reside under those limits. If two family units had been previously identified in the household, the following rules were used to determine the final household person count:

1. When two different family units were already identified in the household, then two different parent sets were being referenced (one of the parent sets was often a single parent). The sum of the two counts (one count might be 0) was used, provided neither pair member had grandparents or grandchildren identified. This was to prevent spouse-spouse pairs from being counted twice, which would happen if grandparents were also parents of children younger than 18 years of age. If two family units were multigenerational families, then the final count was obtained by taking the maximum of the two pair members' counts.

2. It was possible for two different spouse-spouse pairs to be in the household, even though two different family units had not been identified. The final count was set to 2, even though two family units had not been previously identified, under the following conditions:

[8] In rare cases, an individual might identify two spouses in the household. As noted in Section 11.2.2, the true multiplicity count in these cases was not determined; rather, the multiplicity count was set to 1, due to the complexity of determining the appropriate multiplicity count and the rarity of the occurrence of multiple spouses.

- The pair relationship was not a spouse-spouse pair, and the total household size was at least four; and
- Either:
  - Both sides identified a spouse,
  - Both sides identified a partner, or
  - One side identified a parent and the other side identified a parent-in-law.

3. If the conditions for the previous item were not met, it was still possible for two different spouse-spouse pairs to be in the household, even though two different family units were not previously identified. The final count was set to 2 under the following conditions:
   - One pair member had two parents with valid ages and both ages differed from the age of the spouse of the other pair member, and
   - The pair relationship was either sibling-sibling or a pair that was not a pair of interest.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children, grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition—the final count was assigned using the following rules:

4. Among the majority of pairs, the counts for the two sides agreed. However, both sides had to meet the following conditions in order for the final count to be set to one of the sides:
   - The pair could not be a spouse-spouse pair where both respondents had a spouse or both respondents had a partner,
   - No bad relationship codes for roster members aged 15 or older for either pair member,
   - The number of spouse-spouse pairs was either one or zero for both pair members,
   - The household size was greater than 1 and nonmissing on both sides,
   - One pair member had at least two household members aged 15 or older, and
   - There were not two spouse-spouse pairs in the household according to the conditions given in item #3.
5. The counts might have agreed even though the above conditions were not met. The count could still have been set to one of the sides if any one of the following was true:
   - One pair member was younger than 18 and had no bad relationship codes for roster members aged 18 or older, but he or she did have bad relationship codes for

roster members between the ages of 15 and 17 years old. The other pair member had no bad relationship codes for roster members aged 15 or older.

- One pair member had a single bad relationship code, and no other relationship codes could match it to make it a couple (i.e., the pair member did not have a single identified parent, grandparent, parent-in-law, or child-in-law). The other pair member had no bad relationship codes.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where the pair member with no bad roster entries had the same number of household members aged 15 or older as the screener. The pair member with the bad roster entries would not have had the same age composition as the screener if the number of roster members aged 15 or older was added to the number of roster members with bad ages.
- One pair member had bad relationship codes among roster members aged 15 or older or had bad ages, and the other had no bad ages or relationship codes, where all the relationship codes for the pair member with no bad roster entries were immediate family codes (child, parent, sibling, spouse, partner, grandparent, or grandchild). For the pair member with bad roster entries, all the existing relationship codes were immediate family codes.

6. For those cases where the pair was imputed to be a spouse-spouse pair and both sides agreed that only one spouse-spouse pair was in the household, the count was set to 1 if any one of the following conditions was true:
   - Both sides had fewer than four people older than 15 in the household, or
   - One side had fewer than four people older than 15 in the household, and the other side had no bad relationship codes among roster members aged 15 or older
7. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under any one of the following conditions:
   - There were no bad relationship codes among roster members aged 15 or older, or
   - There were no bad relationship codes among roster members aged 18 or older and the pair member had parents.
8. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could have been because a couple entered the household or otherwise materialized after screening. The smaller count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:
   - The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the smaller spouse-spouse count.
   - The screener count of roster members aged 12 or older was smaller than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.

- The difference between the screener count of roster members aged 12 or older and the count of roster members aged 12 or older in the questionnaire rosters of the pair members was smallest with the pair member with the smaller spouse-spouse count.

9. If the count of the number of spouse-spouse pairs did not agree between the two pair members, it could have been because a couple left the household or otherwise dissolved after screening. The larger count was chosen as the final count in this instance, which was identified if the following conditions were satisfied:

   - The screener count of roster members aged 12 or older was no larger than the count of roster members aged 12 or older in the roster of the pair member with the larger spouse-spouse count.
   - The screener count of roster members aged 12 or older was larger than the count of roster members aged 12 or older in the roster of the pair member with the smaller spouse-spouse count.

10. In many cases where the count of the number of spouse-spouse pairs did not agree between the two pair members, one side had a zero count and the other did not. The nonzero count was selected if the pair member associated with the zero count was not a close relative or somehow did not identify a spouse, partner, two parents, or two grandparents. The following conditions were required to select the nonzero count:

    - The pair member with a nonzero count either identified a spouse, a partner, two parents, or two grandparents.
    - The number of roster members aged 15 or older associated with the nonzero count pair member was no larger than the corresponding number associated with the zero count pair member.
    - If the side associated with the nonzero count identified a spouse, partner, or two parents, the following additional conditions were required:
      - The number of roster members between the ages of 26 and 44 was the same between the two pair members.
      - The number of roster members between the ages of 30 and 49 was the same between the two pair members.
      - The number of roster members between the ages of 35 and 54 was the same between the two pair members.
      - The number of roster members between the ages of 40 and 59 was the same between the two pair members.
    - If the side associated with the nonzero count identified two grandparents, the following additional condition was required:
      - The number of roster members aged 50 or older was the same between the two pair members.

11. The counts might not agree because a pair member's partner did not consider the other pair member's family as his or her own family. If at least one side identified a partner and the maximum count was 1, then the maximum was selected if both pair

members had the same number of household members aged 15 or older. Otherwise, if the pair members had a different number of household members aged 15 or older, the count belonging to the pair member with a count of household members aged 15 or older closer to that of the screener was used as the final count.

12. The counts might not agree because a pair member had two grandparents and an uncle/aunt husband-wife pair in the household. The maximum was selected if the pair member associated with the smaller count had a grandparent and had at least two roster members who were neither parents, siblings, children, spouses, partners, or grandparents, and the pair member with the larger count had children-in-law.

13. The count of the number of spouse-spouse pairs might not agree because one of the pairs was a sibling and sibling-in-law, and there are no codes for sibling-in-law. The maximum count was selected if the pair member with the smaller count did not have a spouse or partner but did have siblings aged 15 or older, and there were household members in his or her roster that were not parents, children, siblings, spouses, partners, grandchildren, or grandparents.

14. The count of the number of spouse-spouse pairs might not agree because one side had no nuclear family or grandparent-grandchild relationship codes, and one of the selected respondents was not in a child-parent, child-grandparent, or spouse-spouse relationship. The maximum count was selected if the following conditions were met:

    - The pair member's roster associated with the minimum count (usually 0) had no children, parents, siblings, spouses, partners, grandchildren, or grandparents among respondents aged 12 or older; and
    - The pair member's roster associated with the maximum count had some roster members who were not children, parents, siblings, spouses, partners, grandchildren, or grandparents.

    Note that this condition also nabbed cases where the relationship codes were not correctly identified on one pair member's roster. This occurred rarely, but when it did, the minimum count was 1 and the maximum count was 2.

15. The count of the number of spouse-spouse pairs might not agree because the pair members were siblings, and one sibling did not consider a stepparent or parent's partner as a "parent." The maximum count was selected if the following conditions were met:

    - The pair members were siblings,
    - The pair member associated with the maximum count had two parents,
    - The pair member associated with the minimum count had one parent, and
    - The roster associated with the pair member with the maximum count had more immediate family members (children, parents, siblings, spouses, partners, grandchildren, or grandparents) than the roster associated with the other pair member.

16. The count of the number of spouse-spouse pairs might not agree because the household changed after screening, which was not accounted for by previous conditions. In general, the count with a household composition closest to the screener was selected. The age composition was defined by looking at age classes. The count for a given pair member was selected if any of the following properties held:

    - The number of roster members between the ages of 26 and 44 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
    - The number of roster members between the ages of 30 and 49 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
    - The number of roster members between the ages of 35 and 54 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.
    - The number of roster members between the ages of 40 and 59 for that pair member matched the screener count within the same age range, which differed from the corresponding count for the other pair member.

17. In some cases, neither pair member's household composition matched that of the screener. In that case, the household roster closest to that of the screener was selected. The maximum was selected if the number of screener roster members aged 12 or older exceeded the corresponding count from the questionnaire rosters of both pair members, which also differed from each other.

18. The counts might not agree because, on the rare occasion, one pair member in a spouse-spouse pair identified two grandparents of a different gender. Since there is no code for grandparents-in-law, they could not be identified, so the maximum count was selected. The following conditions were required:

    - The pair was a spouse-spouse pair.
    - The pair member with the maximum count had two grandparents of a different gender, and the pair member with the minimum count did not have any.

    The assumption here, of course, is that the grandparents of a different gender were in fact a spouse-spouse pair. There was no way to check whether a grandfather was the father's father and the grandmother was the mother's mother, for example.

19. Even though the household composition may match in terms of ages across the screener roster and the two pair members' rosters, the counts may disagree where two spouse-spouse pairs were clearly identified by one pair member but not the other. This may be because one of the in-laws was incorrectly identified on one side, or because a partner was not considered an in-law by a responding pair member, or because a partner did not consider other family members as "in-laws." The following conditions were required for the maximum count to be selected:

    - The number of screener roster members aged 12 or older matched the corresponding count from the questionnaire rosters of both pair members.

- The pair member with the maximum number of spouse-spouse pairs had a spouse or partner and also had two parents.
- There were no bad relationship codes among roster members aged 15 or older on either pair member's roster.

20. If the counts for each pair member were not equal but the number of roster members aged 12 or older was the same between the two pair members, and the count for one pair member was the maximum possible in the household, then that number was selected as the final count. This condition was applied only after all other conditions, including conditions where the final count was ambiguous, had already been applied.

21. After accounting for all other rules, if the number of spouse-spouse pairs was still missing, but the lower and upper bounds for imputation were equal to each other, then the final household-level person count was set to one of those bounds.

J.5 Spouse-Spouse Counts (with Children)

The household counts for spouse-spouse counts with children obviously depended upon the counts obtained for spouse-spouse counts with or without children. The first two rules described in this section were determined directly from the spouse-spouse counts or from the household size, and no reconciliation of counts was necessary:

1. For a sizable proportion of cases, clearly no couples with children could be in the household, either because the spouse-spouse count was 0 or the household size was two or less. In these cases, the final spouse-spouse-with-children count was set to 0.

2. An additional small number of cases also could be readily determined by looking at the spouse-spouse count. If one pair member had a spouse-spouse-with-children count that exceeded the final spouse-spouse count, but the other pair member had a spouse-spouse-with-children count that was equal to or smaller than the final spouse-spouse count, then the final spouse-spouse-with-children count was set to the pair member's count that was consistent with the final spouse-spouse count.

The remainder of cases involved households with at least one spouse-spouse couple. After assigning values for the conditions described above, the assignment of values for these cases was done using the rules described in the rest of this section. If two family units had been previously identified in the household, the following rule was used to determine the final household person count:

3. When two different family units were already identified in the household, then two different parent sets were being referenced (one of the parent sets was often a single parent). The sum of the two counts (one count might be 0) was used, provided the spouse-spouse count was greater than 1. In that event, the maximum count was used.

Otherwise, reconciling the counts to a nonmissing value always required the following condition: There was no potential for two or more couples in the household that were not already obviously identified, whereby one of the pair members had at least four roster members of at least 15 years of age. This respondent had grandchildren younger than 18 years of age, did not have children-in-law, and had household members aged 12 or older who were not children,

grandchildren, siblings, children, parents, spouses, or partners. For all remaining cases where a final household count needed to be assigned—in addition to the above condition (unless specifically noted below)—the final count was assigned using the following rules:

4. For cases that were not already determined by looking at the previous two conditions, the counts for the two pair members (if there were two pair members) were equal in the vast majority of cases. The final count could be set to each pair member's count under the following conditions:
    - Both pair members had valid rosters.
    - Either:
        - The counts were nonzero and equal to the final spouse-spouse count, or
        - There were no bad relationship codes for roster members younger than 18, and one of the following conditions held for at least one pair member:
            - The pair member's roster had no bad relationship codes for roster members aged 15 or older,
            - The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or
            - The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 18 in that pair member's roster (covers zero counts since only one bad relationship code could potentially be a single parent but not a pair of parents making a couple).
5. The pair members might both have had zero counts, but the above conditions did not apply. The final count could still have been 0 if the age counts for both pair members and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)
6. The counts for both pair members might still have agreed with nonzero counts, even though none of the previous conditions applied. The final count could still have been set to one of the pair member's counts if the pair relationship was imputed to be a spouse-spouse pair with children.
7. If one pair member did not have a valid roster but the other member did, the final count was set to the other pair member's count under one of the following conditions:
    - The count for the pair member with the valid roster was nonzero and equal to the final spouse-spouse count, or
    - There were no bad relationship codes for roster members younger than 18, and one of the following conditions held for the pair member with the valid roster. Either:
        - The pair member's roster had no bad relationship codes for roster members aged 15 or older,

- The pair member was older than 18 and had neither children nor siblings younger than 18 (covers zero counts since no bad codes were for members younger than 18), or
- The pair member was younger than 18 and did not have parents, but there was one bad relationship code among roster members older than 18 in that pair member's roster (covers zero counts since only one bad relationship code could potentially be a single parent but not a pair of parents making a couple).

8. The pair member with the valid roster might have had a zero count, but the above conditions did not apply. The final count could still have been 0 if the age counts for both the pair member with the valid roster and the screener indicated nobody lived in the household who was younger than 18 and there were no bad roster ages. (In this case, it was not necessary to check for the potential of two or more family units in the household.)

9. If the spouse-spouse-with-children counts disagreed in the same manner as the spouse-spouse counts disagreed, then the choice was obvious: Use the count that corresponded to the correct spouse-spouse count. (In this case, it was not necessary to check for the potential of two or more family units in the household.) Details follow:

   - If the spouse-spouse-with-children counts were equal to the spouse-spouse counts for both pair members, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to the final spouse-spouse count.
   - If the spouse-spouse counts exceeded the spouse-spouse-with-children counts by one for each pair member, even though they were unequal to each other, then the final spouse-spouse-with-children count was set to one less than the final spouse-spouse count.

10. Based on earlier conditions, we already excluded households without couples. We also excluded households with a possibility of two or more couples. If the pair relationship was parent-child and at least one count was nonzero, then the identified couple corresponded to the parent-child relationship. The maximum of the counts was selected under the following conditions:

    - The sum of counts from the two pair members was 1.
    - Either:
      - The relationship was parent-child where the child was between the ages of 12 and 17, or
      - The relationship was parent-child where the child was between the ages of 18 and 20 and the child had siblings younger than 18.

11. In some cases, two couples were identified in the household where the household was multigenerational (one member of the younger couple was in a parent-child relationship with the older couple). If a sibling to the pair member in the younger couple was selected, or if a member of the younger couple was selected who "married into" the family, then he or she was not able to identify the nephews, nieces, brothers-in-law, or sisters-in-law—which could point to an appropriate accounting of all the

couples with children—because of the relationship codes that were available. The maximum of the two counts was selected under the following conditions:

- There were two couples in the household, as identified by the final spouse-spouse count.
- The difference between the pair members' counts was 1.
- Either:
  - The pair member with the smaller count had a spouse or partner and the pair member with the larger count had parents in the household, or
  - The pair member with the smaller count had parents-in-law or children-in-law in the household.

12. If a couple was in a marriage/partnership that occurred after an earlier marriage, the partner might not have considered the partner's children as his or her children, but the child (who also was selected) considered the spouse/partner a parent. Even though the pair relationship was not parent-child, these cases were still counted as spouse-spouse with children since they consisted of the children of one spouse/partner. The maximum count was selected under the following conditions:

    - The pair relationship was not one of interest.
    - One count was 0 and the other count was 1.
    - The pair member with the zero count had a spouse or partner.
    - The pair member with the nonzero count had parents.
    - The spouse-spouse final count was nonmissing.

13. The counts might have been unequal because children younger than 18 left, entered, or otherwise materialized or disappeared in the household after screening and between the time of the interviews. In general, the count was selected that corresponded to the pair member with a household composition closest to the screener household composition. If one pair member did not have children in the household and the other pair member did, the following conditions were required for the count corresponding to the pair member with a household composition closest to the screener:

    - One pair member had a nonzero count of children younger than 18 and the other pair member had a zero count of children younger than 18.
    - Either:
      - The screener composition indicated that children younger than 18 were in the household, whereupon the nonzero count was selected, or
      - The screener composition indicated that no children younger than 18 were in the household, whereupon the zero count was selected.

14. The counts might have been unequal with a count of 0 and a count of 1 because a pair member with a count of 0 was not part of the immediate family unit. The nonzero count was used under the following conditions:

- The pair relationship was not a parent-child, sibling-sibling, spouse-spouse, or grandparent-grandchild relationship.
- Both pair members had relationship codes that were not parent, child, sibling, spouse, partner, grandparent, or grandchild codes among roster members who were aged 12 or older.

15. The counts might have been unequal because of bad relationship codes among roster members younger than 18. The following rules were used to determine if the count associated with the pair member did not have bad relationship codes:
    - The number of roster members younger than 18 was the same between both pair members.
    - The side with the smaller count had at least one bad relationship code for roster members younger than 18.
16. If, after considering all of the general conditions given above, the count was left to imputation, it was still possible that the lower and upper bounds were equal. In this instance, the final count was set to one of the bounds.

Appendix K: Priority Conditions for Creating Household-Consistent Covariates

Appendix K: Priority Conditions for Creating Household-Consistent Covariates

K.1 Household Size

In Table K.1, blank entries indicate that no conditions were required for that set of variables. The reported household size variable is QD54, and the edited household size variable is TOTPEOP, which cannot differ from the raw variable by more than 1. Any variable suffixed by "A" indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to the suffix "B." For example, "QD54A" reflects the reported household size for pair member A. The quality-of-roster counts are considered in the column "any roster missing?" The variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages. The variables that appear in the table are TGOODAGA and TGOODAGB, the total number of cases in the roster with valid ages, incorporating the minimum possible counts within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. Finally, the variable used to describe the screener household size is SHHSIZE. The conditions used to create the variable HHSIZE resulted in no missing values for this variable, and thus no imputation was required. The first column in Table K.1 shows the hierarchical priority condition, with the frequency of occurrence for each priority condition in parentheses.

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
1 (18,317)	Equal, both > 1, both nonmissing	Equal, both > 1, both nonmissing				TOTPEOPA
2 (0)	Equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA ≤ QD54A	A: no		QD54A
3 (0)	Equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGB ≤ QD54B	B: no		QD54B
4 (0)	Equal, both > 1, both nonmissing	TOTPEOPA one more than TOTPEOPB	TGOODAGA = TGOODAGB TGOODAGA ≤ TOTPEOPA		SHHSIZE not equal to QD54A	TOTPEOPA
			TGOODAGA = TOTPEOPA		No condition	
5 (0)	Equal, both > 1, both nonmissing	TOTPEOPB one more than TOTPEOPA	TGOODAGA = TGOODAGB TGOODAGB ≤ TOTPEOPB		SHHSIZE not equal to QD54B	TOTPEOPB
			TGOODAGB = TOTPEOPB		No condition	
6 (0)	Equal, both > 1, both nonmissing	Within one of each other			SHHSIZE at least as large or larger than screener roster, equal to QD54A	SHHSIZE
7 (0)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1, not equal to QD54B	QD54B ≥ TGOODAGB		SHHSIZE ≥ 2, closer to QD54B than TOTPEOPB	QD54B
8 (11)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB ≤ TOTPEOPB (no bad roster ages if equal)		SHHSIZE ≥ 2, TOTPEOPB is as close as QD54B	TOTPEOPB
9 (1)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1	TGOODAGB ≤ SHHSIZE		TGOODAGB ≤ SHHSIZE	SHHSIZE
10 (1)	A: missing or 1 B: not missing > 1	A: missing or 1 B: not missing > 1				TGOODAGB

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
11 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1, not equal to QD54A B: missing or 1	QD54A ≥ TGOODAGA		SHHSIZE ≥ 2, closer to QD54A than TOTPEOPA	QD54A
12 (10)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGA ≤ TOTPEOPA (no bad roster ages if equal)		SHHSIZE ≥ 2, TOTPEOPA is as close as QD54A	TOTPEOPA
13 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1	TGOODAGA ≤ SHHSIZE		TGOODAGA ≤ SHHSIZE	SHHSIZE
14 (0)	A: not missing, > 1 B: missing or 1	A: not missing, > 1 B: missing or 1				TGOODAGA
15 (4)	Both missing or 1	Both missing or 1			SHHSIZE ≥ 2, SHHSIZE at least as large or larger than screener roster	SHHSIZE
16 (35)	Not equal, both > 1	TOTPEOPB = QD54B	A: At least one age range variable less than min.[1] B: Age range variables all same or larger than min.			QD54B
		TOTPEOPA = QD54A	B: At least one age range variable less than min. A: Age range variables all same or larger than min.			QD54A
17 (2)	Not equal, both > 1		A: At least one age range variable less than min. B: At least one age range variable less than min.		Age range variables all same or larger than min.	SHHSIZE

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
18 (923)	Not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54A	QD54A
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: Age range variables all same or larger than min., no bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54B	QD54B
19 (1)	Not equal, both > 1	QD54A is equal to at least one of TOTPEOPA or TOTPEOPB	A: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54A	A: Maxima for each age range between given count and min.
		QD54B is equal to at least one of TOTPEOPA or TOTPEOPB	B: At least one age range variable less than min., or some bad roster ages		SHHSIZE at least as large or larger than screener roster, equal to QD54B	B: Maxima for each age range between given count and min.
20 (2)	Not equal, both > 1	Not equal, both >1	TGOODAGA = TGOODAGB, TGOODAGA = QD54A	A: no B: no		QD54A
			TGOODAGA = TGOODAGB, TGOODAGA = QD54B	A: no B: no		QD54B
21 (0)	Not equal, both > 1	Not equal, both >1	TGOODAGA = QD54A TGOODAGB > QD54B	A: no B: no	SHHSIZE > QD54B	QD54A
			TGOODAGB = QD54B TGOODAGA > QD54A	A: no B: no	SHHSIZE > QD54A	QD54B

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener Roster Characteristics	HHSIZE Equals:
22 (0)	Not equal, both > 1	Not equal, both > 1	TGOODAGA > GOODAGEA, TGOODAGB > GOODAGEB, TGOODAGA = SHHSIZE TGOODAGB = SHHSIZE	A: no B: no	TGOODAGA = SHHSIZE, TGOODAGB = SHHSIZE	SHHSIZE
23 (0)	Not equal, both > 1	Not equal, both > 1	TGOODAGA > GOODAGEA, TGOODAGB > GOODAGEB, TGOODAGA = TGOODAGB	A: no B: no		TGOODAGA
24 (4)	Not equal, both > 1	Not equal, both > 1		A: no B: no	SHHSIZE = sum of maxima for each age group across pair members	SHHSIZE
25 (147)	Not equal, both > 1	Not equal, both > 1		A: no B: no	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to one of the QD54's	QD54A if SHHSIZE closer to A, QD54B if closer to B
					SHHSIZE ≥ 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member

Table K.1 Priority Conditions Used to Create Household-Consistent Household Size (continued)

Priority Condition (Frequency)	Relationship of QD54A & QD54B	Relationship of TOTPEOPA & TOTPEOPB	Relationships Involving Age Range Variables	Any Roster Missing?	Screener HHSIZE Characteristics	HHSIZE Equals:
26 (0)	Not equal, both > 1	Not equal, both > 1		A fewer than B	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to QD54A than QD54B	QD54A
				B fewer than A	SHHSIZE ≥ 2, at least as large or larger than screener roster, closer to QD54B than QD54A	QD54B
				No condition	SHHSIZE ≥ 2, at least as large or larger than screener roster, equidistant between the QD54's	QD54 of oldest pair member
				No condition	SHHSIZE ≥ 2, at least as large or larger than screener roster	SHHSIZE
27 (1)			At least 3 of the age range variables are missing		SHHSIZE ≥ 2, at least as large or larger than screener roster	SHHSIZE

[1] "Min." refers to the minimum possible within each age range based upon the ages of the two pair members.

K.2 Age Variables

Table K.2 illustrates the hierarchical priority conditions ("priorities") used to create a new household-consistent 12 to 17 age group count. Similar priority conditions are used for the 0 to 11, 12 to 14, 12 to 20, 18 to 25, 26 to 34, 35 to 49, 50 or older, and 15 or older age groups. In this table, blank entries indicate that no priority conditions were required for that set of variables. As with the previous table, a variable followed by "A" indicates that the variable corresponds to the value for pair member "A." A similar comment can be made with regard to "B."

As stated earlier, the variables GOODAGEA and GOODAGEB are the total number of cases in the roster with valid ages, and the variables TGOODAGA and TGOODAGB are also the total number of cases in the roster with valid ages, but if the original adjusted count is less than the minimum required, the original count is replaced by the minimum within the age categories 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older. As noted in Section 11.3.1, these counts are adjusted so that the roster ages match what was entered in each pair member's questionnaire. Hence, AGE1217A is the adjusted count of 12- to 17-year-olds for pair member A, and AGE1217B is the adjusted count of 12- to 17-year-olds for pair member B. If AGE1217A or AGE1217B is less than the minimum possible, the count is replaced by the minimum, which is given by TAG1217A and TAG1217B, respectively. Otherwise, AGE1217A and TAG1217A are equivalent, as are AGE1217B and TAG1217B. The sum of AGE011A, AGE1217A, AGE1825A, AGE2634A, AGE3549A, and AGE50PA is GOODAGEA. Similarly, the sum of TAG011A, TAG1217A, TAG1825A, TAG2634A, TAG3549A, and TAG50PA is TGOODAGA. The same can be said for GOODAGEB and TGOODAGB.

The final 12 to 17 age count is denoted by AGE1217. The screener age count, denoted by SAGE1217, is used only if the age counts in each pair member's roster cannot conform to the minimum necessary or otherwise are not possible to incorporate. If after all edits the count for AGE1217 is missing, but the counts for other age groups are not missing, and the counts for the 0 to 11 age group are the same for both pair members, then the sum of the counts for the other age groups, plus the minimum possible for AGE1217, are given by EXC1217. If other means fail to determine the appropriate value for the age count, match measures are used. These are measures that summarize the quality of the match between the two pair members. A match label of "0" indicates a perfect match, where the pair member's roster has a household member who is identified as the other pair member with a perfect match on age and gender and is indicated as the other pair member by the MBRSEL variable. There are several levels of match measures where a lower number signifies a better quality match. These measures are explained in detail in Section 11.2.1.2.1. As a final check, if the age group counts do not equal HHSIZE, and the counts for the pair members are unequal, then the count is set to missing. As with Table K.1, the first column in Table K.2 shows the hierarchical "priority," with the frequency of occurrence for each priority in parentheses, for the AGE1217 count. In most cases, the frequencies corresponding to the other age ranges were the same as the frequency for AGE1217. In those cases where the frequency differed, footnotes provide details of the differences.

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
1 (0)[1]	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB GOODAGEB = HHSIZE, all nonmissing, all > 1	AGE1217A < min. (minimum), AGE1217B ≥ min.				AGE1217B
2 (0)[2]		AGE1217B < min. AGE1217A ≥ min.				AGE1217A
3 (0)		AGE1217A < min. AGE1217B < min.		SHHSIZE = HHSIZE, SAGE1217 ≥ min.		SAGE1217
4 (5)[3]		AGE1217A = AGE1217B, both ≥ min.	Another count except 12-17 < min.			AGE1217A
5 (0)[4]		AGE1217A not equal to AGE1217B, both ≥ min.	AGE1825A < min., AGE1825B ≥ min.			AGE1217B
6 (0)[5]			AGE1825B < min., AGE1825A ≥ min.			AGE1217A
7 (0)			Another count except 12-17 < min.		Fewer roster entries missing in A than B	AGE1217A
8 (0)					Fewer roster entries missing in B than A	AGE1217B
9 (0)[6]					A & B: none missing A has better match measure than B	AGE1217A

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
10 (0)	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, all nonmissing, all > 1	AGE1217A not equal to AGE1217B, both ≥ min.	Another count except 12-17 < min.		A & B: none missing B has better match measure than A	AGE1217B
11 (0)[7]					A & B: none missing Age A ≥ Age B	AGE1217A
					A & B: none missing Age B > Age A	AGE1217B
12 (0)						Missing
13 (17,467)		AGE1217A = AGE1217B	All other counts equal across pair members			AGE1217A
14 (548)		At least one age group has an unequal count between pair members		A: all age counts are equal to their screener counterparts	No missing roster entries on either side	AGE1217A
				B: all age counts are equal to their screener counterparts	No missing roster entries on either side	AGE1217B
15-22 (64)					A & B: none missing A has better match measure than B	AGE1217A
					A & B: none missing B has better match measure than A	AGE1217B
23 (75)					A & B: none missing Age A ≥ Age B	AGE1217A
					A & B: none missing Age B > Age A	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
24 (0)	GOODAGEA = GOODAGEB, GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, all nonmissing, all > 1	At least one age group has an unequal count between pair members			Fewer roster entries missing in A than B A has good match measure (labels 0-7)	AGE1217A
					Fewer roster entries missing in B than A B has good match measure (labels 0-7)	AGE1217B
25 (0)					Fewer roster entries missing in A than B A is older than B	AGE1217A
					Fewer roster entries missing in B than A B is older than A	AGE1217B
26 (0)					Fewer roster entries missing in A than B B is older than A	AGE1217B
					Fewer roster entries missing in B than A A is older than B	AGE1217A
27 (0)					A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
28 (0)					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
29 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A < min. AGE1217B = min.				AGE1217B
30 (0)		AGE1217B < min. AGE1217A = min.				AGE1217A
31 (0)		AGE1217A < min. AGE1217B < min.		SAGE1217 ≥ min.		SAGE1217
32 (0)		AGE1217A = AGE1217B, both ≥ min.				AGE1217A
33 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B ≥ min.			AGE1217B
34 (0)			AGE1825B < min. AGE1825A ≥ min.			AGE1217A
35 (0)					Fewer roster entries missing in A than B	AGE1217A
36 (0)					Fewer roster entries missing in B than A	AGE1217B
37 (0)					A & B: same number of roster entries missing (> 0) A has good match measure (labels 0-7)	AGE1217A
38 (0)					A & B: same number of roster entries missing (> 0) B has good match measure (labels 0-7)	AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
39 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEA = HHSIZE, GOODAGEB not equal to HHSIZE	AGE1217A not equal to AGE1217B			A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B
40 (0)		Priority conditions 29-39 not met				Missing
41 (0)		AGE1217 missing after priority conditions 29-40 invoked, other age range counts not missing				HHSIZE - sum of other age counts
42 (625)		Priority conditions 29-41 not met				AGE1217A
43 (0)[8]	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A < min. AGE1217B = min.				AGE1217B
44 (0)		AGE1217B < min. AGE1217A = min.				AGE1217A
45 (0)[9]		AGE1217A < min. AGE1217B < min.		SAGE1217 ≥ min.		SAGE1217
46 (0)[10]		AGE1217A = AGE1217B, both ≥ min.				AGE1217A
47 (0)		AGE1217A not equal to AGE1217B	AGE1825A < min. AGE1825B ≥ min.			AGE1217B
48 (0)			AGE1825B < min. AGE1825A ≥ min.			AGE1217A

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
49 (0)	GOODAGEA = TOTPEOPA, GOODAGEB = TOTPEOPB, GOODAGEB = HHSIZE, GOODAGEA not equal to HHSIZE	AGE1217A not equal to AGE1217B			Fewer roster entries missing in A than B	AGE1217A
50 (0)					Fewer roster entries missing in B than A	AGE1217B
51 (0)					A & B: same number of roster entries missing (> 0) B has good match measure (labels 0-7)	AGE1217B
52 (0)					A & B: same number of roster entries missing (> 0) A has good match measure (labels 0-7)	AGE1217A
53 (0)[11]					A & B: same number of roster entries missing (> 0) A is older than B	AGE1217A
					A & B: same number of roster entries missing (> 0) B is older than A	AGE1217B
54 (0)		Priority conditions 43-53 not met				Missing
55 (0)		AGE1217 missing after priority conditions 43-54 invoked, other age range counts not missing				HHSIZE - sum of other age counts
56 (442)		Priority conditions 43-55 not met				AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
57 (195)	TGOODAGA = HHSIZE					TAG1217A
	TGOODAGB = HHSIZE					TAG1217B
58 (29)[12]	SHHSIZE = HHSIZE	AGE1217A, AGE1217B ≤ SAGE1217		AGE1217A & B ≤ SAGE1217		SAGE1217
59 (0)	SHHSIZE = HHSIZE, HHSIZE = EXC1217	AGE1217 missing	Other counts not missing, AGE011A equals AGE011B			MIN1217
60 (2)[13]	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, either the two TOTPEOP's > 0, or SHHSIZE = HHSIZE			AGE1217A equals SAGE1217		AGE1217A
				AGE1217B equals SAGE1217		AGE1217B

Table K.2 Priority Conditions Used to Create Household-Consistent Age Variables (Using AGE1217) (continued)

Priority Condition (Frequency)	Relationships Involving TOTPEOP, GOODAGE, and HHSIZE	Relationships Involving AGE1217A, AGE1217B	Relationships Involving Other Age Groups	Relationships Involving Screener Counts	Quality of Roster Measures	AGE1217 Equals:
61 (0)	Previous priority conditions for HHSIZE, TOTPEOP, GOODAGE, not met, SHHSIZE = HHSIZE	AGE1217 missing	At least 3 of the other counts missing			SAGE1217
99 (0)	All prior conditions were not met.					Missing

[1] The following frequencies were observed for priority condition #1: AGE2634 and AGE50p = 1; and AGE011 = 5.

[2] The frequency of priority condition #2 for AGE1825, AGE2634, AGE3549, and AGE50p was 1.

[3] The following frequencies were observed for priority condition #4: AGE011 and AGE3549 = 0; AGE2634 = 1; AGE50p = 2; AGE1825 = 3; and AGE1220=4.

[4] The following frequencies were observed for priority condition #5: AGE1825 = 1 and AGE3549 = 2.

[5] The frequency of priority condition #6 for AGE2634 and AGE3549 was 1.

[6] The frequency of priority condition #9 for AGE2634 and AGE3549 was 1.

[7] The frequency of priority condition #11 for AGE50p and AGE1220 was 1.

[8] The frequency of priority condition #43 for AGE1214 was 1.

[9] The frequency of priority condition #45 for AGE1217 and AGE1220 was 1.

[10] The frequency of priority condition #46 for AGE1825, AGE2634, and AGE50p was 1.

[11] The frequency of priority condition #53 for AGE15p and AGE3549 was 1.

[12] The following frequencies were observed for priority condition #58: AGE011 = 23; AGE1825 and AGE1220 = 28; AGE3549 = 30; and AGE50p = 32.

[13] The following frequencies were observed for priority condition #60: AGE011 = 1; and AGE1217, AGE3549, AGE1214, and AGE1220 = 3.

Appendix L: Creation of Household-Level and Person-Level Files

Appendix L: Creation of Household-Level and Person-Level Files

L.1 Introduction

For the 2012 administration of the National Survey on Drug Use and Health (NSDUH), a person was randomly selected for an interview through a four-stage sample selection process. States were first stratified into a total of 900 State sampling (SS) regions. Within each of these SS regions, a sample of census tracts was selected (i.e., the first stage of selection) with probabilities proportional to a composite size measure and with minimum replacement. Within sampled census tracts, adjacent census blocks were combined to form the second-stage sampling units or area segments. One area segment was selected within each sampled census tract with probability proportional to population size.[1] Once the sample segments were selected, specially trained field staff visited areas and created lists of all eligible dwelling units (DUs) within the sample segment boundaries. These lists served as the frames for the third stage of sample selection. After the DUs were selected within each segment, an interviewer visited each selected DU to obtain a roster of all persons aged 12 or older. This roster information was then used to select zero, one, or two persons from the household at the fourth stage of sample selection.

At the end of the survey year, a household-level file and a person-level file were created to record the information obtained from the sampling processes described above. The household-level and person-level files were used in the final creation of the person-level and pair-level analysis weights. In addition, the person-level file was later subset into a smaller data file that contained only respondents who were considered "completed" cases; this file was used for analysis. Refer to Section L.3 for the definition of a completed case.

L.2 Dwelling Unit-Level Eligibility and Completeness Criteria

Before proceeding with the fourth stage of sample selection (i.e., within selected households), a set of rules was used to determine whether or not a DU was eligible to be selected. Eligibility of the DU was recorded in the binary variable DUELIG, where a value of 1 indicated eligibility. Two examples of DUs deemed to be ineligible included those defined as "vacant" and those determined to be "not a primary residence."

Occasionally, DUs were eligible but failed to complete the screening process. Reasons for not completing the screening process were recorded, including situations such as "language barrier," "refusal," and "denied access." Completeness of the screening process for the DU was recorded in the binary variable DUCOMP, where a value of 1 indicated completeness. For the segments where all the DUs were from denied-access areas, such as gated communities, an adjustment was made in the final household-level file. Although the field interviewers could not obtain an accurate count of DUs from denied-access areas, these DUs were considered eligible.

[1] Segments consist of clusters of the geographic aggregated adjacent census blocks. SS regions were formed through geographically partitioning each State into roughly equal-sized regions based on a composite size measure. The 2012 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013) contains more information regarding the sample design.

Therefore, DU information from the U.S. Census Bureau for these areas was used in the household-level file.

During the second stage of sampling, it was possible to select a sample segment more than once because samples were selected with replacement. These duplicated segments had different segment IDs (SEGIDs) for each duplicate. However, one SEGID contained all the DU information and the other had none. The number of eligible DUs was split as evenly as possible between the two SEGIDs, and this information was updated in the household-level and person-level files.

L.3 Person-Level Eligibility and Completeness Criteria

During screening, respondents were asked to identify all eligible household members so that only eligible individuals were listed and, therefore, potentially selected. Eligibility was determined according to the criteria provided in Section L.1. Eligible respondents at the time of screening were recorded in the binary variable PRELIG, which had a value of 1 if the household member was eligible. Respondents who were selected were recorded in the binary variable PRSEL, where 1 indicated a selected individual. It was possible for an individual to be selected, but at the time of the interview, to be determined ineligible. Examples of changes from eligibility to ineligibility included "the selected person turned out not to be a permanent resident in the DU" and "roster error." If this occurred, the value of PRELIG was changed from 1 to 0.

A summary of the number of selected, eligible, and completed dwelling units is shown in Table L.1. The number of eligible, selected, and interviewed persons also is summarized in the table.

Table L.1 NSDUH Household, Person Eligibility, and Completed Interview Counts: 2012

	Selected Dwelling Units	Eligible Dwelling Units	Completed Screenings (Dwelling Units)	Eligible Persons	Selected Persons	Inter-viewed Persons	Completed Cases (Interviews)
CAI	214,274	178,586	153,873	325,641	87,656	68,360	68,309

CAI = computer-assisted interviewing.

To be considered a completed case for purposes of analysis, a respondent had to provide "yes" or "no" answers to the cigarette usage gate question and to at least 9 of the following additional drug usage gate questions: (1) chewing tobacco, (2) snuff, (3) cigars, (4) alcohol, (5) marijuana, (6) cocaine (in any form), (7) heroin, (8) hallucinogens, (9) inhalants, (10) pain relievers, (11) tranquilizers, (12) stimulants, and (13) sedatives.[2] Unlike the paper-and-pencil interviewing (PAPI) questionnaire in 1999 and surveys prior to 1999, no logical inference could be made from information within a section if the gate question was not answered. This was because the computer-assisted interviewing (CAI) instrument routed respondents out of a section if the gate question was not answered. Completeness of the survey for eligible individuals was recorded in the binary variable PRCOMP, which had a value of 1 if the respondent was a completed case, and 0 if not. For a summary of the number of completed cases in the 2012 survey, see Table L.1.

[2] For more details on the "usable case rule," see Section 2.2.3.

L.4 Variables in the Household-Level and Person-Level Files

This section documents some of the important person-level variables that were created for the household-level and person-level files.

Screener-level demographic variables were created from the screener roster information in the household-level and person-level files. XAGE was the screener age, which either could be "continuous" (single-year ages) or categorical. A respondent could choose to give an age category instead of the actual age. The age categories with their accompanying codes were 199 = 12 to 17 years old; 299 = 18 to 25 years old; 399 = 26 to 34 years old; 499 = 35 to 49 years old; and 599 = 50 years old or older. Screener race (XRACE1-XRACE6), screener Hispanicity (XHISP), and screener gender (XSEX) also were produced from the screener roster information. XRACE1 through XRACE6 were indicator variables representing white, black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, and other, respectively. The household-level variable PAIRSEL represented the number of persons within each age group selected from a DU. It was a 20-level variable indicating whether zero, one, or two individuals were selected from the five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older) in a given household. (If two persons were selected from the household, this variable indicated the age groups of both pair members.) Similar to PAIRSEL, the household-level variable PAIRRESP had 20 levels, which indicated whether zero, one, or two persons completed the interviews from the five age groups within a household.

As described in the 2012 NSDUH sample design report (Morton et al., 2013), States were partitioned into SS regions, which were further partitioned into clusters of adjacent blocks called "segments." The variable SEGID (segment ID number) was a two-letter State abbreviation followed by a two-digit SS region and a two-digit segment identifier, which uniquely identified each segment. Census region (REGION) was a four-level geographic variable recoded from the respondent's State of residence. The four levels were Northeast, Midwest, South, and West. The population density variables PDEN00 and PDEN200[3] classified respondents according to their living situation, whether it was in a rural or urban area, and, if urban, the size of the urban area. They were used to categorize segments where the respondents lived according to the modified 2000 census data, which was adjusted to more recent data from Claritas, Inc.[4] PDEN200 had five levels: segment in core-based statistical area (CBSA)[5] with 1 million or more persons; segment in CBSA with 250,000 to 999,999 persons; segment in CBSA with fewer than 250,000 persons; segment in urban area but not in CBSA; and segment in rural area (not in CBSA and not in urban area). PDEN00 had three levels: segment in CBSA with 1 million or more persons; segment in CBSA with fewer than 1 million persons; and segment not in CBSA.

The variable PLACNM00[6] was the census place name associated with each segment. According to the census documentation, this variable was defined as places, for the reporting of

[3] PDEN00 and PDEN200 were previously named PDEN and PDEN2. The suffix "00" indicates that the variables were derived from the 2000 census.

[4] Claritas, Inc., is a market research firm headquartered in San Diego, California.

[5] CBSAs, developed in response to standards put forth by the Office of Management and Budget (OMB) in June 2003, are metropolitan and micropolitan areas that were designated using data from the 2000 census. More information about CBSAs can be retrieved from http://www.census.gov/hhes/www/housing/resseg/cbsa.html.

[6] PLACNM00 was previously named PLACNAME. The suffix "00" indicates that the variables were derived from the 2000 census.

decennial census data, which includes census-designated places, consolidated cities, and incorporated places. If duplicate place names existed within the same county, the places were distinguished by their legal description (e.g., "city" or "village"). However, because the variable PLACNM00 was used to help the field interviewers locate the segment and was limited by the number of characters printed on the map, identifiers like "city" or "village" have been removed from the place name. The variable STATE represented the Federal Information Processing Standards (FIPS) State codes for the 50 States and the District of Columbia. The variable STATE was created at the sampling stage and did not contain any missing values.

The variables VESTR and VEREP were created to capture the sampling design structure. Each SS region appeared in a different variance estimation stratum (VESTR) every quarter. Two replicates (VEREP) were defined within each variance stratum. Each replicate consisted of four segments, one for each quarter of data collection. The segment-level variable RURORURB is derived from the 2000 census block-level designation of rural or urban.[7] In the NSDUH sample, if 100 percent of the blocks are rural, the segment is defined as rural (RURURB00[8] = 1). If one or more of the blocks within a segment is urban, the segment is defined as urban (RURURB00 = 2). Other sampling variables such as DIVISION, SSREGION, GQTYPE, ID, STNAME, STUSAB, and QUARTER[9] also were included in the household-level and person-level files.

[7] The census classifies as urban all blocks located within urbanized areas (UA) and urban clusters (UC). UAs and UCs generally consist of core census block groups or blocks that have a population density of at least 1,000 persons per square mile and surrounding census blocks that have an overall density of at least 500 persons per square mile. In addition, under certain conditions, less densely settled territory may be part of each UA or UC.

[8] RURURB00 was previously named RURORURB. The suffix "00" indicates that the variables were derived from the 2000 census.

[9] For more details on these sampling variables, refer to the 2012 NSDUH sample design report (Morton et al., 2013).

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

PERSON-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.107.004
Phase I, Deliverable No. 41

Authors:

Patrick Chen
Devon Cribb
Lanting Dai
Harper Gordek
Jeff Laufenberg
Neeraja Sathe
Matthew Westlake

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2014

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

PERSON-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.107.004
Phase I, Deliverable No. 41

Authors:

Patrick Chen
Devon Cribb
Lanting Dai
Harper Gordek
Jeff Laufenberg
Neeraja Sathe
Matthew Westlake

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2014

Preface

This report contains a brief review of the sampling weight calibration methodology used for the 2012 National Survey on Drug Use and Health (NSDUH), which was known as the National Household Survey on Drug Abuse (NHSDA) prior to 2002. This report also lists detailed documentation on the implementation steps and evaluation results from the weight calibration application. The constrained exponential modeling (CEM) method used in the surveys prior to 1999 (referred to in this report as the generalized exponential model [GEM]) was modified to provide more flexibility in dealing internally with the extreme weights and for setting bounds directly on the weight adjustment factors so they can become suitable for nonresponse (nr) and poststratification (ps) adjustments. The highlights of the method are summarized below.

- The inherent two-phase nature of the NSDUH design (viewing the large screener sample as the first phase and the actual questionnaire sample as the second phase) allows for the additional step of poststratifying the selected persons to estimated controls from the large first-phase sample of persons. This additional step results in stable controls for the later step of nonresponse adjustment at the respondent-person level. These two steps had been combined as one step in surveys prior to 1999, but they have been kept separate from 1999 onward.
- A poststratification step at the respondent-household level in the first phase of the screening interview reduced coverage bias resulting from the first-phase sampling and produced controls for use in poststratification at the selected-person level, respondent person-pair level, and respondent-household level in the second phase of the drug use interview. This step again takes advantage of the inherent two-phase design of the study.
- The built-in control on extreme weights in GEM was supplemented by a separate step of extreme value adjustment after the final poststratification whenever the extreme weight percentage in the initial unadjusted weights was considered to be too large. This was accomplished by using GEM so that the sample demographic distribution was preserved. This method represents an improvement over the trimming method implemented before the nonresponse adjustment in surveys prior to 1999 and the extreme value adjustment before the nonresponse adjustment used for the 1999 NHSDA. For the 2012 NSDUH, this final extreme value adjustment was judged to be unnecessary.

The GEM calibration method provides a unified approach to handling problems of extreme weights, nonresponse, and poststratification, and it uses current state-of-the-art technology. The implementation of GEM under a tight project schedule was a challenge, but it was met successfully by the diligence and perseverance of the members of the weighting team consisting of Patrick Chen, Devon Cribb, Lanting Dai, Harper Gordek, Jeff Laufenberg, Neeraja Sathe, and Matthew Westlake.

This report consists of several chapters describing the implementation and evaluation of GEM and of appendices composed mainly of tables. In the interest of reducing the size of the report, detailed domain-specific evaluation results are presented in the supplement to this report, which is available upon request. This work was completed for the Substance Abuse and Mental

Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International,[1] North Carolina, under Contract No. HHSS283201000003C. The authors are grateful to Art Hughes of SAMHSA for his useful comments and suggestions.

Ralph Folsom, Senior Advisor
Research Triangle Park, NC

[1] RTI International is a trade name of Research Triangle Institute.

Table of Contents

Chapter **Page**

1. Introduction 1
2. Generalized Exponential Model for Weight Calibration 5
3. Predictor Variables in GEM for the 2012 NSDUH 7
4. Practical Aspects of Implementing GEM for the NSDUH 11
   - 4.1 Definition of Extreme Weights of Sampling Weights 11
   - 4.2 Definition of Lower and Upper Bounds for Weight Adjustment Factors 11
   - 4.3 Definition of Control Totals 13
   - 4.4 Efficient Computation Using Grouped Data 13
   - 4.5 Steps in GEM Fitting 14
   - 4.6 Quality Control Checks 14
   - 4.7 Practical Guidelines in Using GEM 15
   - 4.8 Variable Collapsing Guide 17
5. Weight Calibration at Phase I Dwelling Unit and Phase II Person Levels 19
   - 5.1 Phase I Household-Level Weight Components 22
     - 5.1.1 Weight Components #1 to #7: Selection of a Dwelling Unit 22
     - 5.1.2 Weight Component #8: Dwelling Unit–Level Nonresponse Adjustment 23
     - 5.1.3 Weight Component #9: Dwelling Unit–Level Poststratification Adjustment 23
     - 5.1.4 Weight Component #10: Dwelling Unit–Level Extreme Weight Adjustment 24
   - 5.2 Phase II Person-Level Weight Components 25
     - 5.2.1 Weight Component #11: Selection of a Person within a Dwelling Unit 25
     - 5.2.2 Weight Component #12: Selected Person-Level Poststratification Adjustment 26
     - 5.2.3 Weight Component #13: Respondent Person-Level Nonresponse Adjustment 26
     - 5.2.4 Weight Component #14: Respondent Person-Level Poststratification Adjustment 26
     - 5.2.5 Weight Component #15: Respondent Person-Level Extreme Weight Adjustment 27
6. Evaluation of Calibration Weights 29
   - 6.1 Response Rates 29
   - 6.2 Percentages of Extreme Weights and Outwinsors 29
   - 6.3 Slippage Rates 30
   - 6.4 Weight Adjustment Summary Statistics 31
   - 6.5 Sensitivity Analysis of Drug Use Estimates to Baseline Models 31

References 45

Appendix **Page**

A Technical Details about the Generalized Exponential Model A-1

B Poststratification Control Totals B-1

C Imputation Methodology C-1

D Generalized Exponential Model Summary D-1

- D1 Model Group 1: New England (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont) D-17
- D2 Model Group 2: Middle Atlantic (New Jersey, New York, and Pennsylvania) D-29
- D3 Model Group 3: East North Central (Illinois, Indiana, Michigan, Ohio, and Wisconsin) D-41
- D4 Model Group 4: West North Central (Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota) D-53
- D5 Model Group 5: South Atlantic (Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia) D-65
- D6 Model Group 6: East South Central (Alabama, Kentucky, Mississippi, and Tennessee) D-77
- D7 Model Group 7: West South Central (Arkansas, Louisiana, Oklahoma, and Texas) D-89
- D8 Model Group 8: Mountain (Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming) D-101
- D9 Model Group 9: Pacific (Alaska, California, Hawaii, Oregon, and Washington) D-113

E Evaluation of Calibration Weights: Response Rates E-1

F Evaluation of Calibration Weights: Dwelling Unit–Level Percentages of Extreme Weights and Outwinsors F-1

G Evaluation of Calibration Weights: Person-Level Percentages of Extreme Weights and Outwinsors G-1

H Evaluation of Calibration Weights: Slippage Rates H-1

I Evaluation of Calibration Weights: Weight Summary Statistics I-1

List of Tables

Table **Page**

5.1 Sample Size, by Model Group for Each Stage of Sampling 21

5.2 Weight Distribution for Design-Based Weight and Weight after DU-Level Adjustments 25

5.3 Weight Distribution for Weight before Any Person-Level Adjustment and after Person-Level Adjustments 27

6.1 Summary Statistics of Overall Weighted Response Rates across Individual States 29

6.2 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH 33

6.3 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH 35

6.4 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH 37

6.5 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH 39

6.6 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH 41

6.7 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH 43

D Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (United States) D-4

D.1a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 1: New England) D-19

D.1b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 1: New England) D-20

List of Tables (continued)

Table **Page**

D.2a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 2: Middle Atlantic) D-31

D.2b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 2: Middle Atlantic) D-32

D.3a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 3: East North Central) D-43

D.3b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 3: East North Central) D-44

D.4a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 4: West North Central) D-55

D.4b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 4: West North Central) D-56

D.5a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 5: South Atlantic) D-67

D.5b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 5: South Atlantic) D-68

D.6a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 6: East South Central) D-79

D.6b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 6: East South Central) D-80

D.7a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 7: West South Central) D-91

D.7b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 7: West South Central) D-92

D.8a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 8: Mountain) D-103

D.8b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 8: Mountain) D-104

D.9a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 9: Pacific) D-115

D.9b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 9: Pacific) D-116

E.1 2012 NSDUH Weighted Response Rates: United States, District of Columbia, and the 50 States E-3

F.1 2012 NSDUH Dwelling Unit–Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States F-3

List of Tables (continued)

Table **Page**

G.1 2012 NSDUH Selected Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States G-3

G.2 2012 NSDUH Respondent Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States G-5

H.1 2012 NSDUH Slippage Rates: UNITED STATES H-3

H.2 2012 NSDUH Slippage Rates: ALABAMA H-3

H.3 2012 NSDUH Slippage Rates: ALASKA H-4

H.4 2012 NSDUH Slippage Rates: ARIZONA H-4

H.5 2012 NSDUH Slippage Rates: ARKANSAS H-5

H.6 2012 NSDUH Slippage Rates: CALIFORNIA H-5

H.7 2012 NSDUH Slippage Rates: COLORADO H-6

H.8 2012 NSDUH Slippage Rates: CONNECTICUT H-6

H.9 2012 NSDUH Slippage Rates: DELAWARE H-7

H.10 2012 NSDUH Slippage Rates: DISTRICT OF COLUMBIA H-7

H.11 2012 NSDUH Slippage Rates: FLORIDA H-8

H.12 2012 NSDUH Slippage Rates: GEORGIA H-8

H.13 2012 NSDUH Slippage Rates: HAWAII H-9

H.14 2012 NSDUH Slippage Rates: IDAHO H-9

H.15 2012 NSDUH Slippage Rates: ILLINOIS H-10

H.16 2012 NSDUH Slippage Rates: INDIANA H-10

H.17 2012 NSDUH Slippage Rates: IOWA H-11

H.18 2012 NSDUH Slippage Rates: KANSAS H-11

H.19 2012 NSDUH Slippage Rates: KENTUCKY H-12

H.20 2012 NSDUH Slippage Rates: LOUISIANA H-12

H.21 2012 NSDUH Slippage Rates: MAINE H-13

H.22 2012 NSDUH Slippage Rates: MARYLAND H-13

H.23 2012 NSDUH Slippage Rates: MASSACHUSETTS H-14

H.24 2012 NSDUH Slippage Rates: MICHIGAN H-14

H.25 2012 NSDUH Slippage Rates: MINNESOTA H-15

H.26 2012 NSDUH Slippage Rates: MISSISSIPPI H-15

H.27 2012 NSDUH Slippage Rates: MISSOURI H-16

List of Tables (continued)

Table **Page**

H.28 2012 NSDUH Slippage Rates: MONTANA H-16
H.29 2012 NSDUH Slippage Rates: NEBRASKA H-17
H.30 2012 NSDUH Slippage Rates: NEVADA H-17
H.31 2012 NSDUH Slippage Rates: NEW HAMPSHIRE H-18
H.32 2012 NSDUH Slippage Rates: NEW JERSEY H-18
H.33 2012 NSDUH Slippage Rates: NEW MEXICO H-19
H.34 2012 NSDUH Slippage Rates: NEW YORK H-19
H.35 2012 NSDUH Slippage Rates: NORTH CAROLINA H-20
H.36 2012 NSDUH Slippage Rates: NORTH DAKOTA H-20
H.37 2012 NSDUH Slippage Rates: OHIO H-21
H.38 2012 NSDUH Slippage Rates: OKLAHOMA H-21
H.39 2012 NSDUH Slippage Rates: OREGON H-22
H.40 2012 NSDUH Slippage Rates: PENNSYLVANIA H-22
H.41 2012 NSDUH Slippage Rates: RHODE ISLAND H-23
H.42 2012 NSDUH Slippage Rates: SOUTH CAROLINA H-23
H.43 2012 NSDUH Slippage Rates: SOUTH DAKOTA H-24
H.44 2012 NSDUH Slippage Rates: TENNESSEE H-24
H.45 2012 NSDUH Slippage Rates: TEXAS H-25
H.46 2012 NSDUH Slippage Rates: UTAH H-25
H.47 2012 NSDUH Slippage Rates: VERMONT H-26
H.48 2012 NSDUH Slippage Rates: VIRGINIA H-26
H.49 2012 NSDUH Slippage Rates: WASHINGTON H-27
H.50 2012 NSDUH Slippage Rates: WEST VIRGINIA H-27
H.51 2012 NSDUH Slippage Rates: WISCONSIN H-28
H.52 2012 NSDUH Slippage Rates: WYOMING H-28
I.1 2012 NSDUH Dwelling Unit–Level Weight Summary Statistics: United States, District of Columbia, and the 50 States I-3
I.2 2012 NSDUH Selected Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States I-5
I.3 2012 NSDUH Respondent Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States I-7

List of Exhibits

Exhibit **Page**

1.1 Sampling Weight Calibration Steps 3

3.1 Definition of Levels for Variables 8

4.1 Generalized Exponential Model Steps 12

5.1 Summary of 2012 NSDUH Sample Weight Components 20

5.2 U.S. Census Bureau Divisions/Model Groups 21

5.3 Imputed Demographic Variables and Corresponding Explanatory or Auxiliary Sort Variables 24

B.1 Definition of Levels for Variables B-3

D.1 Definition of Levels for Variables D-6

D.2 Covariates for 2012 NSDUH Person Weights (res.sdu.nr) D-12

D.3 Covariates for 2012 NSDUH Person Weights (res.sdu.ps) D-13

D.4 Covariates for 2012 NSDUH Person Weights (sel.per.ps and res.per.nr) D-14

D.5 Covariates for 2012 NSDUH Person Weights (res.per.ps and res.per.ev) D-15

D1.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 1: New England D-23

D1.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 1: New England D-24

D1.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 1: New England D-25

D1.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 1: New England D-26

D1.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 1: New England D-27

D2.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 2: Middle Atlantic D-35

D2.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 2: Middle Atlantic D-36

D2.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 2: Middle Atlantic D-37

D2.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 2: Middle Atlantic D-38

D2.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 2: Middle Atlantic D-39

List of Exhibits (continued)

Exhibit **Page**

D3.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 3: East North Central D-47

D3.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 3: East North Central D-48

D3.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 3: East North Central D-49

D3.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 3: East North Central D-50

D3.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 3: East North Central D-51

D4.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 4: West North Central D-59

D4.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 4: West North Central D-60

D4.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 4: West North Central D-61

D4.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 4: West North Central D-62

D4.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 4: West North Central D-63

D5.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 5: South Atlantic D-71

D5.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 5: South Atlantic D-72

D5.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 5: South Atlantic D-73

D5.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 5: South Atlantic D-74

D5.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 5: South Atlantic D-75

D6.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 6: East South Central D-83

D6.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 6: East South Central D-84

D6.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 6: East South Central D-85

List of Exhibits (continued)

Exhibit **Page**

D6.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 6: East South Central D-86

D6.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 6: East South Central D-87

D7.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 7: West South Central D-95

D7.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 7: West South Central D-96

D7.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 7: West South Central D-97

D7.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 7: West South Central D-98

D7.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 7: West South Central D-99

D8.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 8: Mountain D-107

D8.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 8: Mountain D-108

D8.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 8: Mountain D-109

D8.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 8: Mountain D-110

D8.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 8: Mountain D-111

D9.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 9: Pacific D-119

D9.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 9: Pacific D-120

D9.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 9: Pacific D-121

D9.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 9: Pacific D-122

D9.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 9: Pacific D-123

List of Terms and Abbreviations

C	Center point.
CAI	Computer-assisted interviewing.
DU	Dwelling unit.
ev	Extreme weight adjustment. See Section 4.1 for more detail.
FI	Field interviewer.
GEM	Generalized exponential model. See Chapter 2 for more detail.
half-step	This refers to halving the increment in the Newton-Raphson iterative process for fitting GEM.
IQR	Interquartile range.
L	Lower bound on adjustment factor.
MPMN	Multivariate predictive mean neighbor.
nr	Nonresponse adjustment.
Outwinsor	Signifies the percentages of weights trimmed after extreme weight adjustment via winsorization.
PMN	Predictive mean neighborhood.
ps	Poststratification adjustment.
res.sdu.nr	Respondent screener dwelling unit nonresponse adjustment step. See Section 5.1.2 for more detail.
res.sdu.ps	Respondent screener dwelling unit poststratification adjustment step. See Section 5.1.3 for more detail.
res.sdu.ev	Respondent screener dwelling unit extreme weight adjustment step. See Section 5.1.4 for more detail.
sel.per.ps	Selected person poststratification adjustment step. See Section 5.2.2 for more detail.
res.per.nr	Respondent person nonresponse adjustment step. See Section 5.2.3 for more detail.
res.per.ps	Respondent person poststratification adjustment step. See Section 5.2.4 for more detail.
res.per.ev	Respondent person extreme weight adjustment step. See Section 5.2.5 for more detail.
SAE	Small area estimate.
SDU	Screener dwelling unit.
SE	Standard error.
SES	Socioeconomic status indicator. See Exhibit 3.1 for more detail.
SS	State sampling.
U	Upper bound on adjustment factor.
UPMN	Univariate predictive mean neighbor.
UWE	Unequal weighting effect. It refers to the contribution in the design effect due to unequal selection probability and is defined as $1 + [(n - 1)/n]*CV^2$ where CV = coefficient of variation of weights, and n is the sample size.
VESTR	Variance estimation stratum.
VEREP	Variance estimation replicates.
Winsorization	A method of extreme weight adjustment that replaces extreme weights with the critical values used for defining low and high extreme weights.

1. Introduction

The target population for the 2012 National Survey on Drug Use and Health (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing within the United States. A coordinated sample design was developed for the 2005 through 2009 NSDUHs. The 2010 through 2011 and 2012 through 2013 samples are two extensions of the 5-year sample. Although there is no planned overlap with the 1999 to 2004 samples, the coordinated design for 2005 through 2009 facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. This design was intended to reduce field costs and increase the precision of estimates in year-to-year trend analyses, using the expected positive correlation resulting from the overlapping sample between successive NSDUH years. The 2012 NSDUH main sample continues the 50 percent overlap by retaining half of the second-stage units from the 2011 survey.

The 2012 design provides for estimates by State in all 50 States plus the District of Columbia. States may therefore be viewed as the first level of stratification as well as a reporting variable. Eight States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas), referred to as the "large" States, had a sample designed to yield 3,600 respondents per State, while the remaining 43 "small" States (which include the District of Columbia) had a sample designed to yield 900 respondents per State. Sample in these 43 States supports reliable State estimates based on small area estimation (SAE) or direct estimation methodology when several years of data are combined. The target national sample size for the 2012 NSDUH was 67,500 persons, and the achieved sample for the 2012 NSDUH was 68,309 persons—corresponding to 48,850 responding dwelling units [DUs] selected at the second phase out of 153,858[2] DUs screened at the first phase, in which the first phase is screening and the second phase is interview. The achieved sample has a low of 829 for Arkansas to a high of 976 for West Virginia among small States, and a low of 3,544 for Florida to a high of 3,687 for Ohio among large States.

In the 2012 NSDUH design, States served as the primary strata; within each State, State sampling (SS) regions were formed and served as the secondary strata. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. The smaller States were partitioned into 12 SS regions, whereas the 8 large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions.

Unlike previous NSDUHs, the first stage of selection for the 2005 through 2012 NSDUHs was census tracts selected from SS regions. This stage was included to contain sample segments within a single census tract to the extent possible. Prior to the 2005 NSDUH, segments that crossed census tract boundaries made merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were

[2] The number of DUs that completed the first-phase screening was 153,873, but some DUs did not have eligible persons, so they were removed from DU poststratification and person-level calibration steps. The number of DUs that had eligible persons in them was 153,858.

aggregated within SS regions until each tract had, at a minimum, 150 DUs in urban areas and 100 DUs in rural areas. There were 48 census tracts per SS region selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979).

Because census tracts generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census tract. For this second stage of sampling, each selected census tract was partitioned into compact clusters[3] of DUs by aggregating adjacent census blocks. Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." A sample DU in NSDUH refers to either a housing unit or a group-quarters listing unit, such as a dormitory room or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement will support the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

One segment was selected within each sampled census tract with probability proportionate to size. The 48 selected segments then were randomly assigned to a survey year and quarter of data collection.

After sample segments for the 2012 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the third stage of sample selection. Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame.

After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. Using the roster information obtained from an eligible member of the selected DU, zero, one, or two persons were selected for the survey. Sampling rates were preset by age group and State. Roster information was entered directly into the electronic screening instrument, which automatically implemented this fourth stage of selection based on the State and age group sampling parameters.

As in previous years of the survey,[4] the 2012 NSDUH sample weighting posed challenges because of the sheer magnitude of the number of State-specific predictors used for nonresponse (nr) and poststratification (ps) adjustments. With the 51-State survey, using a single model for each of the adjustments was not practical; however, treating each State separately was not desirable because individual State sample sizes were not large enough to support reliable estimation of a number of parameters. Therefore, the 51 States were grouped into nine model groups corresponding to the nine U.S. Census Bureau divisions. This helped to keep a substantial number of predictor variables in each model and reduced the computing time that would be associated with fitting a larger model.

---

[3] Although the entire cluster is compact, the final sample of DUs represents a noncompact cluster. Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. Although compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish, 1965).

[4] The survey was known as the National Household Surveys on Drug Abuse (NHSDA) prior to 2002.

As with each survey after 1999, an important feature of the 2012 NSDUH sample weighting was to capitalize on the inherent two-phase nature of the NSDUH design (although the design was primarily viewed as multistage) by adding a step to poststratify the household weights in the first phase of the screening interview (see Exhibit 1.1). This reduced coverage bias resulting from the first phase of sampling and produced estimated controls for use in poststratification of person-pair weights and household weights in the second phase of the drug use interview. No other suitable source was available for obtaining these controls for poststratification. Note also that screener DU weights were poststratified to population counts by adjusting the DU's weighted contribution of person counts to various demographic domains. The second important feature was to add a step to poststratify selected persons (including respondents and nonrespondents) to estimated controls from the large first-phase sample of persons for various predictor variables at the segment, DU, and person levels. This provided stable controls for the step involving the nonresponse adjustment of respondent weights. Incorporating this important feature would not have been possible without screener data on the sociodemographics of members of the selected households.

Exhibit 1.1 Sampling Weight Calibration Steps

As in previous NSDUHs, a modification of the earlier methodology of scaled constrained exponential modeling (CEM) (Folsom & Witt, 1994) was used to meet the new demands on the weighting mentioned previously (i.e., the two-phase design and large number of available predictors). The modified methodology, called the generalized exponential model (GEM) (Folsom & Singh, 2000), has several features:

- Like CEM, GEM can use a large number of predictor variables, such as those obtained from the first-phase screener sample for the 50 States plus the District of Columbia, and some of their interactions.
- GEM allows unit-specific bounds for the weights initially identified as extreme, which provide tight controls on the extreme weights. This built-in control is often adequate, in that the frequency of extreme weights, after the nonresponse and poststratification adjustments, is not usually high. However, if this is not the case, GEM can be used for a separate extreme weight adjustment after poststratification. This extra adjustment, which uses tighter bounds, will preserve the demographic population controls used in the poststratification step.
- GEM provides a unified approach to nonresponse, poststratification, and extreme weight adjustments. The differences are only in terms of the bounds and control totals that are used.
- GEM can be implemented efficiently using software developed at RTI.
- GEM is a generalization of the commonly used raking-ratio method in which a distance function is minimized such that (1) the initial weights are perturbed only a little and lie within certain bounds, and (2) control totals are met. It is also a generalization of Deville and Särndal's (1992) logit method in that the bounds on weights are not required to be uniform. Moreover, the lower bound can be set to one, which is desirable for the nonresponse adjustment. Like the previously mentioned methods, fitting GEM requires iterations (such as Newton-Raphson).

The report is organized as follows. In Chapter 2, GEM is reviewed, and a heuristic description outlines how GEM provides a unified approach to all three procedures' adjustments for nonresponse, poststratification, and extreme weight adjustment. In Chapter 3, potential predictor variables for use with nonresponse, poststratification, and extreme weight are discussed, and the strategy for dealing with many predictors via modeling groups of States is reviewed. In Chapter 4, practical steps for implementing GEM for the 2012 NSDUH are presented, and in Chapter 5, details of the weight calibrations, including all weight components corresponding to Phases I and II, are given. Chapter 6 presents the evaluation measures of calibrated weights and a sensitivity analysis of point estimates and standard errors (adjusted for calibration) of selected drug prevalence estimates. The sensitivity analysis compares the estimates and standard errors from final models to those of the baseline models (which consist of only main effects). Nine appendices also are included. Appendix A presents technical details about GEM; Appendix B documents the creation and source of the poststratification control totals; and Appendix C contains information on the imputation methodology. Appendix D summarizes the GEM modeling, and the remaining five appendices contain various tables on weighted response rates, percentages of extreme weights and outwinsors, slippage rates, and weight adjustment summary statistics.

2. Generalized Exponential Model for Weight Calibration

In survey practice, design weights are typically adjusted in three steps via the following methods: (1) weighting class adjustments for nonresponse, (2) raking-ratio adjustments for poststratification, and (3) winsorization for extreme weights. The bias introduced by winsorization is alleviated to some extent through poststratification. The nonresponse (nr) adjustment is a correction for bias that is introduced when estimates are based only on responding units; poststratification is an adjustment for coverage (typically undercoverage) bias, as well as for variance reduction (which is possibly due to correlation between the study and control, usually demographic, variables). If weights are not treated for extreme weight adjustment, the resulting estimates, although unbiased, will tend to have low precision.

There are limitations in the existing methods of weight adjustment for nonresponse, poststratification, and extreme weight. For the nonresponse step, there are general raking-type methods, such as the scaled constrained exponential model developed by Folsom and Witt (1994), where the lower and upper bounds can be suitably chosen by using a separate scaling factor. The factor is set as the inverse of the overall response propensity. It would be beneficial to have a model for the nonresponse adjustment factor that incorporates the desired lower and upper bounds on the factor as part of the model. Note that the lower bound on the nonresponse adjustment factor should be 1 because it is interpreted as the inverse of the probability of response for a particular unit. For the poststratification step, the general calibration methods of Deville and Särndal (1992), such as the logit method, allow for built-in lower (L) and upper (U) bounds (for poststratification, typically L < 1 < U). However, it would be useful to have nonuniform bounds $(\mathrm{L}_k, \mathrm{U}_k)$ depending on the unit *k*, such that the final adjusted weights, w_k, could be controlled within certain limits. An important application of this feature would be weight adjustments to allow the user to have some control over the final adjustment of weights initially identified as extreme weights. It would be advantageous to adjust for bias introduced in the extreme weight adjustment step (such as when extreme weights are treated via winsorization) so that the sample distribution for various demographic characteristics is preserved.

A modification of the earlier method of the scaled constrained exponential model of Folsom and Witt (1994), termed the generalized exponential model (GEM) and proposed by Folsom and Singh (2000), provides a unified approach to the three weight adjustments for nonresponse, poststratification, and extreme weight, and it has the valuable features mentioned previously. The functional form of the GEM adjustment factor is given in Appendix A. It generalizes the logit model of Deville and Särndal (1992), typically used for poststratification, such that the bounds (L, U) may depend on *k*. Thus, it provides a built-in control on extreme weights, during both nonresponse adjustments and poststratification. In addition, the bounds are internal to the model and can be set to chosen values (e.g., $\mathrm{L}_k = 1$ in the nonresponse step). If the frequency of extreme weights is low after the final poststratification, a separate extreme weight adjustment step may not be necessary.

Note that in view of the nonresponse adjustment factor being defined as the inverse of response propensity, GEM requires it to be greater than 1. However, the built-in extreme weight

control feature of GEM essentially defines adjustment factors with regard to the critical value under winsorization. Therefore, although the adjustment factor with regard to the cutoff point is always greater than 1, with regard to the original weight, it can be less than 1.

In fitting GEM to a particular problem, choosing a large number of predictor variables along with tight bounds will have an impact on the resulting unequal weighting effect (UWE) and the percentage of extreme weights. In practice, this leads to somewhat subjective evaluations of trade-offs between the target set of bounds for a given set of factor effects, the target UWE, and the target proportions of extreme weights. The percentage of "outwinsors" (a term coined to signify the extent of residual weights after extreme weight adjustment via winsorization) is probably a more realistic benchmark in determining the robustness of estimates in the presence of extreme weights. Chapter 4 provides details about the GEM process and some practical guidelines about fitting such a model. In particular, an adaptive method based on realized minimum and maximum bounds after setting loose initial bounds is recommended for choosing bounds more objectively.

A large increase in the number of predictor variables in GEM typically would result in a higher UWE, indicating a possible loss in precision. By looking at the change in variance calculated for a model run with the minimal number of predictor variables versus the final model we reached during the weighting process, a more precise measure of loss (or gain) in precision can be obtained for variance of selected study variables. The results are presented in Chapter 6.

3. Predictor Variables in GEM for the 2012 NSDUH

For the 2012 National Survey on Drug Use and Health (NSDUH), the initial set of predictor variables was identical to the set used for the 2011 NSDUH. Exhibit 3.1 shows the definitions and levels of these predictor variables. Typical predictors used for the screener dwelling unit (DU) nonresponse adjustment were State, Quarter, Group-Quarters Indicator, Population Density, Percentage Hispanic or Latino in Segment, Percentage Black or African American in Segment, Percentage Owner-Occupied DUs in Segment, and Segment-Combined Median Rent and Housing Value, which is also called the Socioeconomic Status (SES) indicator. The SES indicator was a composite measure based on (standardized) median rent, median housing value, and the percentage of dwellings that are owner occupied. Typical predictors for the person-level nonresponse adjustments were, in addition to those stated previously, Age, Gender, Race, Hispanicity, and Relation to Householder (i.e., the head of the household). For poststratification, predictors typically used were State, Age, Race, Gender, Hispanicity, and Quarter. In all cases, the model consisted of main effects and some interactions of these predictors. For a separate extreme weight adjustment with the generalized exponential model (GEM) after poststratification, the predictors were the same as those used in the poststratification (ps) adjustment.

Generally, it is desirable to include, whenever possible, poststratification predictors (correlated with the outcome variable) as part of nonresponse predictors (correlated with the response variable) because of the potential variance reduction; this works to offset the variance inflation, which is due to the random controls used in the nonresponse (nr) adjustment. In general, this is not possible because demographic information (often used for poststratification) is not available for nonrespondents. However, with a two-phase design, such as NSDUH's, this problem does not exist because the screener data contain the necessary information. There is, of course, the cost in time and effort required to edit and impute the screener-based predictors in advance of this nonresponse adjustment. Many times, the need to edit, impute, or both edit and impute nonresponse predictors for the full sample, which consists of respondents and nonrespondents, is eliminated because the poststratification and nonresponse adjustments are combined into a single poststratification step. However, the processes leading to nonresponse and coverage errors are likely to be different enough to benefit from separate modeling. The nonresponse-adjustment models also can benefit from bias reduction when segment-level variables, such as the percentage of owner-occupied DUs, are included in the model. Population totals for these segment-level variables have not been developed for use as poststratification controls.

Exhibit 3.1 Definition of Levels for Variables

Age (years)
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1,4]

Gender
1: Male, 2: Female[1]

Group Quarters Indicator
1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Hispanicity
1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]

Percent of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)
1: 50% - 100%,[1] 2: 10% - 50%, 3: <10%

Percent of Segments That Are Black or African American
1: 50% - 100%, 2: 10% - 50%, 3: <10%[1]

Percent of Segments That Are Hispanic or Latino
1: 50% - 100%, 2: 10% - 50%, 3: <10%[1]

Population Density
1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race (3 level)
1: White,[1] 2: Black or African American, 3: Other

Race (5 level)
1: White,[1] 2: Black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: Two or More Races

Relation to Householder
1: Householder or Spouse,[1] 2: Child, 3: Other Relative, 4: Nonrelative

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

States[3]

Model Group 1:	1: Connecticut, 2: Maine, 3: New Hampshire, 4: Rhode Island, 5: Vermont, 6: Massachusetts[1]
Model Group 2:	1: New Jersey,[1] 2: New York, 3: Pennsylvania
Model Group 3:	1: Illinois, 2: Indiana,[1] 3: Michigan, 4: Wisconsin, 5: Ohio
Model Group 4:	1: Iowa, 2: Kansas, 3: Minnesota, 4: Missouri,[1] 5: Nebraska, 6: South Dakota, 7: North Dakota
Model Group 5:	1: Delaware, 2: District of Columbia, 3: Georgia,[1] 4: Maryland, 5: North Carolina, 6: South Carolina, 7: Virginia, 8: West Virginia, 9: Florida
Model Group 6:	1: Alabama, 2: Kentucky, 3: Mississippi, 4: Tennessee[1]
Model Group 7:	1: Arkansas,[1] 2: Louisiana, 3: Oklahoma, 4: Texas
Model Group 8:	1: Colorado, 2: Idaho, 3: Montana, 4: Nevada, 5: New Mexico, 6: Utah, 7: Wyoming, 8: Arizona[1]
Model Group 9:	1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington,[1] 5: California

MSA = metropolitan statistical area.

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-Combined Median Rent and Housing Value (also known as the Socioeconomic Status [SES] indicator) is a composite measure based on rent, housing value, and percent owner occupied.

[3] The States assigned to a particular model are based on census divisions.

[4] The age group 50+ was further broken down into 50-64 and 65+ for Person-Level Poststratification Adjustment, for which 65+ was used as the reference level.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Heuristically, the suitable number of State-specific controls should depend on the size of the realized sample in each State; because of this, the nature of the problem of too many controls in nonresponse- and poststratification-adjustment models is State specific. Therefore, for the 2012 NSDUH, the strategy proposed by Singh, Penne, and Gordek (1999) was followed and is discussed in the following paragraphs. Also using Singh et al. (1999), some general guidelines were used to choose an initial set of State-specific controls, and the initial set was modified iteratively as problems in maintaining them arose. The process began with the baseline model of one-factor effects and then proceeded with the addition of second- and third-order effects; collapsing was performed as necessary, depending on the individual State sample sizes. To obtain more precise State-level estimates, every effort was made to include as many important State-specific covariates as possible in models for nonresponse and poststratification weight adjustments. These covariates typically were defined by sociodemographic domains. However, keeping a multitude of State-specific covariates, especially higher order interactions, was not possible because individual State sample sizes were not large enough to support stable estimation of an adequate number of model parameters. Therefore, a hierarchical order was used for including covariates in the model; the order started with covariates at the national level, followed by covariates at the census division level within the Nation, then covariates at the combined State level within the census division, and finally, whenever possible, covariates at the State level within the combined States.

When adding certain covariates to the model resulted in parameters that could not be estimated or were unstable, the hierarchy strategy mentioned previously was used to combine States within a census division so that covariates at the combined level could be included. However, this problem typically arose with State-specific higher order interactions, and States were collapsed only when combining levels of covariates within a State was not a reasonable alternative. This was thought to be beneficial in obtaining more reliable State-level estimates using small area estimation (SAE) techniques. The eight large States were not combined with other smaller States, to the extent possible, so that direct State-level estimates could be obtained without relying on SAE.

As an objective check for the suitability of the number of factors, once a satisfactory convergent model was obtained (see Section 6.5 for details), the relative efficiency of a more complex model (with many effects) versus a simpler model (with fewer effects) was measured. In addition to the relative efficiency, the increase in the unequal weighting effect (UWE) was checked.

For the 2012 NSDUH data, as for the previous years' data, it became apparent that the number of controls could be very high (in excess of 1,000). This many controls would be computationally prohibitive because the implementation of GEM involves iterative steps, and a matrix (whose dimension corresponds to the number of controls) must be inverted in each of these iterations. A solution would be to use separate models within groups of States rather than a single overall model. It can be shown that, if effects (two-factor or higher order) are always collapsed within a group of States, then fitting an overall model of GEM is equivalent to fitting separate models for each group. In this way, the computational problems associated with too many controls could be reduced. Therefore, in the 2012 NSDUH, as in the 1999 through 2011 surveys, nine model groups corresponding to the nine census divisions were used.

4. Practical Aspects of Implementing GEM for the NSDUH

As explained in Chapter 2, the generalized exponential model (GEM) can be used for nonresponse (nr) adjustment, poststratification (ps), and extreme weight adjustment (see Exhibit 4.1 for a schematic presentation of the steps). These steps were implemented using the GEM macro developed at RTI. A detailed discussion can be found in Chen, Penne, and Singh (2000).

4.1 Definition of Extreme Weights of Sampling Weights

An important aspect of GEM is the built-in provision of extreme weight adjustment. Sampling weights for the survey generally were classified as extreme (high or low) if they fell outside the commonly used interval defined by the median $\pm 3 \times$ interquartile range (IQR) for some prespecified domains; these domains were usually defined by design strata, taking into account deep stratification. For example, the dwelling unit (DU)-level weight for the 2012 National Survey on Drug Use and Health (NSDUH) used the State sampling (SS) region as the domain. The person-level weight adjustments used a hierarchy of four domains: (1) SS region × Age group, (2) State × Age group, (3) SS region, and (4) State. A minimum of 30 observations was required for defining the boundaries, or critical values, for extreme weights. If this minimum was not met at the lower level, the next level up in the hierarchy was used.

Although the SS region × Age group domain corresponded to a deep stratum, it could be unsuitable for defining extreme weights because of insufficient sample sizes. So, collapsing SS regions within a State gave rise to such domains as State × Age group. Even at this level, sample sizes could be insufficient, so SS regions and, later, States themselves could be used as domains to define extreme weights. The critical values for low and high extreme weights are denoted by $b_{k(l)}$ and $b_{k(u)}$, respectively. The critical points for extreme weights within GEM modeling were defined as the median $\pm 2.5 \times$ IQR, which was conservative when compared with the commonly used standard of the median $\pm 3 \times$ IQR. This is because, to better prevent the adjusted weights from crossing the standard boundary and those at or beyond the boundary, weights near but below it (which have the most potential to become extreme) were treated as extreme by GEM.

4.2 Definition of Lower and Upper Bounds for Weight Adjustment Factors

For implementing extreme weight control via GEM, the variable m_k was defined as $b_{k(u)} / w_k$ for high extreme weights, and $b_{k(1)} / w_k$ for low extreme weights, where w_k represents the sampling weight before adjustment, and $b_{k(u)}, b_{k(l)}$ denote the critical values for the extreme weights. (Note that under this definition, nonextreme weights has a value of 1 for m_k; for high extreme weights, the more extreme the weight is, the smaller m_k will be; conversely for low extreme weights, the more extreme the weight is, the bigger m_k will be.)

Exhibit 4.1 Generalized Exponential Model Steps

GEM = generalized exponential model; SE = standard error; UWE = unequal weighting effect.

The upper and lower bounds for the adjustment factors were defined, respectively, as the product of m_k and the upper and lower boundary parameters specified in the modeling of GEM.

GEM allows inputs of three different upper (U) and lower (L) boundary parameters (L_1, and U_1, L_2, and U_2, L_3, and U_3, respectively) for high, non-, and low extreme weights. By applying a small upper boundary parameter for high extreme weights and a large lower boundary parameter for low extreme weights, the extreme weights could be controlled in the modeling.

GEM also requires specification of centers (C), such that L < C < U. For nonresponse adjustment, it was constructive to require all adjustments to be greater than 1 because the adjustments represented the inverse of response propensities. The value of C in this case was chosen as the inverse of the overall response propensity. For poststratification, centers were set to 1 so the adjusted weights would not be too far away from the original design weights. Here, lower bounds were chosen to be less than 1 and upper bounds were greater than 1 because the control totals could be larger or smaller than the estimated totals based on the design weights. The extreme weight adjustment is analogous to the poststratification adjustment (see Appendix A) in that it is a repeated poststratification with tighter bounds for extreme weights identified after the poststratification step. Section 4.7 gives guidelines for the choice of lower, center, and upper parameters.

4.3 Definition of Control Totals

GEM modeling for nonresponse adjustment, poststratification, and extreme weight adjustment involved estimation of parameters of the adjustment factor model, such that specified control totals were satisfied. There were two types of control totals. For nonresponse adjustment, the control totals were from the full sample (i.e., respondents and nonrespondents), while for poststratification, control totals were obtained from external sources, such as the Census Bureau or a large first-phase screener sample. Specifically, for the 2012 NSDUH, the control totals for various domains for the selected person-level poststratification adjustment (sel.per.ps, see Section 5.2.2) were obtained from the first-phase sample containing roster information, and the control totals for the respondent person-level poststratification (res.per.ps, see Section 5.2.4) were obtained from the Census Bureau's Postcensal Population Estimates for various demographic domains. Controls used for extreme weight adjustment were the same as those for poststratification because they were based on the poststratified weight. (See Appendix B for more information.)

4.4 Efficient Computation Using Grouped Data

Because adjustment factors remained the same for units (DUs or persons) having common values for all explanatory variables used in the model, the size of the sample data was reduced by grouping units having common values of these variables. Also, within the groupings, the units with extreme weights were further grouped such that, in addition to the common values of the explanatory variables, they also had common values of m_k. This significantly saved computation time, especially because the original sample size was large. Modeling GEM with grouped data was implemented by treating each group as a single record, with the associated weight defined as the sum of the individual weights in the group. Note that when using GEM with grouped data, the unequal weighting effect (UWE) and *t*-test statistics normally produced in

the output would be misleading because the weights in grouped data are sums of the weights for the individual units within each group. Also, the definition of variance estimation stratum (VESTR) and replicates (VEREP) required for variance calculation would not be correct. To avoid these misleading results from using the grouped data, the final model was rerun with the full (ungrouped) data.

4.5 Steps in GEM Fitting

Exhibit 4.1 depicts the GEM steps. After specifying the GEM parameters, such as the initial upper and lower bounds, the number of the Newton-Raphson iterations and half-steps, and the type of weight adjustment (nonresponse adjustment, poststratification, or extreme weight adjustment), a forward selection method for modeling was used. A model with only main effects and loose bounds was first fit to obtain a set of realized baseline upper and lower bounds for extreme and nonextreme weights and to calculate a baseline UWE. Next, using the realized bounds, as many higher order interactions as possible were added to the model to help reduce bias, without unduly increasing the UWE and the extreme weight percentages. Convergence problems were addressed by loosening lower bounds and upper bounds and collapsing or dropping variables. In GEM, *t* tests and *p* values for significance of various effects could be computed for a previously converged model, which would be helpful in deciding about the collapsing of effects when convergence problems arose with realized bounds.

For this application, "collapsing" implies combining the "levels" of variables with other levels explicitly present in the model, while "dropping" implies combining with the reference levels, which are not explicitly represented in the model. Collapsing or dropping lower order interactions had a direct impact on the inclusion of the number of higher order interactions. For the 2012 NSDUH, when adding higher order terms, all previously selected explanatory variables were retained in the model. Possible reasons for nonconvergence included explanatory variables corresponding to domains with small sample sizes, or domains with large discrepancies between estimated totals based on the initial weights and the target control totals. The variables causing problems with convergence were identified by the high magnitude of the estimated model parameters. Once the explanatory variables were finalized, finer adjustments of upper bounds and lower bounds could optimize the model by reducing UWE and the extreme weight percentages.

4.6 Quality Control Checks

The distributions of the weights before and after each adjustment were compared to uncover any unusual impact of the weight adjustment on the initial weights. In addition to the weight distributions, the ratios of the maximum weight to the mean weight and the UWEs were compared across various domains both before and after each adjustment. The percentages of extreme weights were checked after each adjustment to see how effective the modeling was in controlling extreme weights. Coverage bias analysis based on the slippage rates also was conducted to check the impact of poststratification on various noncontrolled domains (i.e., those factors that were dropped or collapsed in the model).

4.7 Practical Guidelines in Using GEM

1. Collapsing checks for domains with small sample sizes. The number of observations in various domains defined by levels of the factor effects was examined. If the domain sample size was 0 and the control total corresponding to this domain also was 0, the factor generally was dropped. This automatically collapsed the factor level with the reference level; however, if the control total was not 0, the factor could not be dropped because collapsing the domains together for the sample also would collapse the population domains together. The result would be that control totals could not be met for the reference levels involved. In these cases, the factor level corresponding to a 0 domain sample size should be collapsed with another level for which we are willing to compromise on satisfying the control total.

In general, domains with small sample sizes may cause problems during GEM modeling and prevent the model from converging. For the 2012 NSDUH, if the model did not converge because a domain sample size was small, the corresponding factor effect was collapsed with another effect based on substantive considerations. For example, if State was involved, then it was better, in general, to collapse within States; collapsing of geographically adjacent States was done only when there was no other reasonable alternative (see Section 4.8 for more details). The necessity of collapsing was checked at each stage of model enlargement in the forward selection of factors. If variables were collapsed at a previous stage, the corresponding factor levels were also collapsed using the hierarchy principle at succeeding stages involving higher order factor effects.

2. Singularity checks. As in the case of collapsing checks, singularity checks (i.e., linear dependence checks of realized value columns of the predictors) were performed for the baseline model; in addition, they were performed at each stage of model enlargement because singularities depended on what other predictors were in the model. (Note that, although all variables were linearly independent of each other, it was possible for the columns of their realized values to have been linearly dependent.) For nonresponse adjustment, any variable that was a linear combination of other variables was either dropped from the model or collapsed with other variables. To decide whether to drop or to collapse, a singularity check was performed for both respondents only and the full sample. If both samples showed the same set of variables causing singularity, then these singularity variables could be dropped; if not, collapsing needed to be performed. For poststratification adjustment, any variable that was a linear combination of other variables had to be collapsed with other variables because the variables corresponding to poststratification controls typically were linearly independent.

3. Finding the initial factor set. After the collapsing and singularity checks, the remaining factor effects at a given stage of model enlargement formed the initial factor set.

4. Baseline model. Starting with the model consisting of all one-factor effects from the initial factor set, a convergent version was found (after any required collapsing) under no restrictions on the bounds. The model was optimized by trying to reduce the UWE and tighten the bounds. If necessary (to obtain convergence), factors corresponding to large parameter estimates were collapsed. As an option, p values could have been used to determine which factors to collapse.

5. Baseline plus two-factor effects. All two-factor interactions from the initial factor set were added to the baseline model. A convergent version under no bound restrictions then was found, and the model was optimized using criteria described in Guideline 4. The non-State two-factor effects were added first, and then, in a separate step, the State two-factor effects were added.

6. Baseline with two and higher order factor effects. Starting with the optimized model from Guideline 5, the higher order factor effects were added—first the non-State three-factor effects, then, in a separate step, the State three-factor effects. Again, criteria from Guideline 4 were followed to obtain an optimal model.

7. Optimizing a model with respect to the target model characteristics. These are summarized in the following points:

- For each step of model enlargement, the UWE for the initial weights was computed. It was allowed to increase up to 20 percent, or the maximum allowable UWE (generally under six), whichever was lower.
- The following guidelines, based on empirical considerations, were used for setting the bounds. In the case of poststratification and separate extreme weight adjustments, the center was set as $C_1 = C_2 = C_3 = 1$. Instead of tightening the bounds to as close to 1 as possible, as was done for surveys prior to 2002, we used an adaptive approach to choose the bounds starting from the 2003 NSDUH; that is, starting with loose bounds of (0.1, 10), we performed GEM iteratively four times, each with the realized bounds from the previous iteration. The final bounds for nonextreme weights were desired to be around (0.2, 5). The iterations based on the adaptive approach generally met this desired criterion. If this was not the case, then collapsing of some model variables was allowed to meet this criterion. Finally, the bounds U_1 and L_3 were further tightened to be as close to 1 as possible to better control high and low extreme weights, while maintaining $L_3 \geq L_2$ and $U_1 \leq U_2$.
- In the case of nonresponse, the centers were set equal to the common value of the overall inverse response propensity, and all the three lower bounds (L_1, L_2, and L_3) were set to 1. Next, starting with the loose bounds of (1, 10), the bounds were chosen iteratively as mentioned above using the realized bounds from the previous GEM iteration. The bounds U_1 and L_3 were further tightened to as close to center as possible, while maintaining $L_3 \geq L_2$ and $U_1 \leq U_2$.
- Targets for the maximum acceptable percentages of extreme weights and outwinsors within GEM for nonresponse and poststratification were as follows: 3 percent for the unweighted extreme weights, 15 percent for weighted extreme weights, and 5 percent for outwinsors. These percentages are liberal and serve as guidelines only. In practice, reducing them by half is preferable. If these guidelines were not met after all stages of calibration, a separate GEM for adjustment of extreme weights was implemented after poststratification.

8. Evaluation measures. After each stage of model enlargement, various characteristics were examined for large values. These included the UWE, the ratio of the maximum to the mean

for adjusted weight, the percentage of extreme weights and outwinsors, the distance between the total sample weighted count and the target population count (i.e., slippage rates for different domains), and other characteristics, such as weight summary statistics. In addition, the distributions of adjustment factors were checked for highly asymmetric tails. With the set of realized bounds for the final model, the baseline model was rerun, and then point estimates and SEs for selected outcome variables for the two models were compared. Generally, the two estimates were likely to be close, but not the SEs. The SEs for the final model were expected to be smaller but, at times, could be larger. Larger SEs were identified and examined because they could be an indication of instability of the model parameter estimates because of possible overfitting or insufficient sample sizes. In such situations, the final model was revised to get a more parsimonious model.

4.8 Variable Collapsing Guide

As discussed in Section 4.5, convergence problems in GEM were solved by either loosening bounds or collapsing model variables. Grouping proposed levels into a smaller number of categories could be done in several ways, but care was taken so that they remained meaningful. When constructing the model and attempting to obtain convergence, maintenance of logical groupings was a top priority. The following are some general guidelines that were followed when collapsing variables.

- *Ordinal variables*. Most of the proposed explanatory variables were ordinal. Thus, collapsing was done in a meaningful way, following the order. For example, the combined rental/house quintile had five levels (i.e., 1st, 2nd, 3rd, 4th, and 5th quintile) with the 5th quintile set for the reference. If the 4th quintile needed to be collapsed, it would be collapsed with either the 3rd or 5th quintile.
- *Age groups*. Age group had five levels: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older (50 or older was further broken down into 50 to 64 and 65 or older for the person-level poststratification adjustment and the person-level extreme weight adjustment to increase the accuracy of estimates for these age groups). For the main effects, the age covariate with five or six levels was easy to incorporate in the model. For the interactions, every effort was made to maintain the age group, and, therefore, collapsing was performed within age groups first. Collapsing across age groups occurred only if the age groups could not be maintained separately.
- *Large and adjacent States*. In the main effects, fitting States separately in the model was not a problem. For the State-specific interactions, collapsing was done within the State first, collapsing with other adjacent States was done only if needed. For the eight States with large sample sizes (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, Texas), every effort was made to preserve all factor levels within States so that direct estimates could be made for the large States.
- *Race*. In the main effects and State-specific two-factor interactions, Race had five levels (white, black or African American, American Indian or Alaska Native, Asian, and two or more races), while in non-State-specific two- and three-factor effects, Race had three levels (white, black or African American, and other). If maintaining all five levels was difficult in the main effects or State × Race interactions, the following guidelines were followed: (1) collapse American Indian or Alaska Native

and Asian if either of them caused a convergence problem; (2) collapse black or African American with two or more races if black or African American caused a convergence problem; (3) collapse two or more races with American Indian or Alaska Native or Asian, whichever had a smaller sample size, if two or more races caused a convergence problem; and (4) collapse American Indian or Alaska Native, Asian, and two or more races, or collapse all nonwhite Race groups if necessary. In the State × Race interactions, collapsing Race was done within State. If the three-level Race could not be maintained, the levels were collapsed to white and nonwhite.

5. Weight Calibration at Phase I Dwelling Unit and Phase II Person Levels

The 2012 National Survey on Drug Use and Health (NSDUH) was based on probability sampling so that valid inferences could be made from survey findings to the target population. Probability sampling refers to sampling in which every unit on the frame is given a known, nonzero probability of inclusion in the survey. This is required for unbiased estimation of the population total. The assumption of nonzero inclusion probability for every pair of units in the frame also is required for unbiased variance estimation. The basic sampling plan involved four stages of selection across two phases of design (see Exhibit 5.1). The first phase of the design was the dwelling unit (DU) level, and the second phase was the person level. The four stages of selection were as follows: within Phase I, (1) the selection of census tracts within the State sampling (SS) region; (2) the selection of segments within each sampled census tract; (3) the selection of DUs within these segments; and within Phase II, (4) the selection of eligible individuals within DUs (Table 5.1). Specific details of the sample design and sample selection procedures can be found in the 2012 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

As part of the postsurvey data-processing activities, analysis weights were calculated for the 2012 NSDUH respondents that reflected the selection probabilities from various stages of the sample design. These sample weights were adjusted at both the DU level (screening sample) and person level (drug questionnaire sample) to account for bias due to extreme weights, nonresponse, and coverage.

The final Phase I DU-level and Phase II person-level sample weights for the 2012 NSDUH sample are products of several factors (see Exhibit 5.1), each representing either a probability of selection at some particular stage or some form of extreme weight, nonresponse, or poststratification adjustment. In the following sections, these components are described in greater detail. In summary, the first 10 factors are defined for all screener-complete DUs and reflect the fully adjusted DU-level weight. The latter five components reflect the person-level selection within each screened DU, as well as any additional adjustments for person-level extreme weight, nonresponse, and poststratification error. Note that the unconditional, final person-level weights for the 2012 NSDUH sample are the product of all 15 weight components, as illustrated in Exhibit 5.1.

Exhibit 5.2 shows the U.S. Census Bureau divisions and model groups used in the 2012 NSDUH person-level weight calibration.

Exhibit 5.1 Summary of 2012 NSDUH Sample Weight Components

Phase I Dwelling Unit Level

Design Weight Components	
#1	Inverse Probability of Selecting Census Tract
#2	Inverse Probability of Selecting Segment
#3	Quarter Segment Weight Adjustment
#4	Subsegmentation Inflation Adjustment
#5	Inverse Probability of Selecting Dwelling Unit
#6	Inverse Probability of Added/Subsampled Dwelling Unit
#7	Dwelling Unit Release Adjustment

Weight Adjustment Components	
#8	Dwelling Unit Nonresponse Adjustment ***(res.sdu.nr)****
#9	Dwelling Unit Poststratification Adjustment ***(res.sdu.ps)****
#10	Dwelling Unit Extreme Weight Adjustment ***(res.sdu.ev)****

Phase II Person Level

Design Weight Components	
#11	Inverse Probability of Selecting a Person within a Dwelling Unit

Weight Adjustment Components	
#12	Selected Person-Level Poststratification Adjustment to Screener Data Controls ***(sel.per.ps)****
#13	Person-Level Nonresponse Adjustment ***(res.per.nr)****
#14	Person-Level Poststratification Adjustment ***(res.per.ps)****
#15	Person-Level Extreme Weight Adjustment ***(res.per.ev)****

* These adjustments use the generalized exponential model (GEM), which also involves pre- and postprocessing in addition to running the GEM macro. See Exhibit 4.1. For computational feasibility, all weight adjustments were done using the nine model groups based on U.S. Census divisions defined in Exhibit 5.2.

Exhibit 5.2 U.S. Census Bureau Divisions/Model Groups

Model Group	Census Division
1	**New England (6 States)**
	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
2	**Middle Atlantic (3 States)**
	New Jersey, New York, Pennsylvania
3	**East North Central (5 States)**
	Illinois, Indiana, Michigan, Ohio, Wisconsin
4	**West North Central (7 States)**
	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota
5	**South Atlantic (8 States and the District of Columbia)**
	Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia
6	**East South Central (4 States)**
	Alabama, Kentucky, Mississippi, Tennessee
7	**West South Central (4 States)**
	Arkansas, Louisiana, Oklahoma, Texas
8	**Mountain (8 States)**
	Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming
9	**Pacific (5 States)**
	Alaska, California, Hawaii, Oregon, Washington

Table 5.1 Sample Size, by Model Group for Each Stage of Sampling

Model Group	Eligible DU	Completed DU	Eligible Persons	Selected Persons	Completed Persons
1	15,380	13,365	27,964	7,174	5,615
2	25,030	19,503	42,272	11,127	8,158
3	33,444	28,338	59,924	16,573	12,800
4	15,937	14,672	29,581	7,926	6,342
5	30,681	26,239	54,883	13,632	10,883
6	8,961	8,223	16,782	4,534	3,656
7	14,468	13,032	27,530	8,113	6,347
8	17,185	15,671	32,856	9,054	7,282
9	17,500	14,830	33,849	9,523	7,226
Total	178,586	153,873	325,641	87,656	68,309

DU = dwelling unit.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

In the 2012 NSDUH, as in the 2000 through 2011 surveys, the order of the extreme weight adjustment step at both the DU and person level was different from the order used in the 1999 National Household Survey on Drug Abuse (NHSDA) computer-assisted interviewing (CAI). In the 1999 NHSDA CAI, the extreme weight adjustment step was introduced before nonresponse and poststratification, which was analogous to the traditional trimming step before nonresponse and poststratification. In the 1999 NHSDA, the initially identified extreme weights were held fixed at their winsorized values, and the nonextreme weights were adjusted so that the original sample distribution of the weights for various domains was preserved. As a better alternative for the surveys after 1999, the generalized exponential model (GEM) first was allowed to control the extreme weights during the nonresponse and poststratification steps, and then a separate extreme weight adjustment step was performed after poststratification, if necessary. This step would be like a repeated poststratification, except that the extreme weights identified after poststratification would have tighter bounds, thus preserving the sample distributions in various domains (equivalent to satisfying the poststratification controls). For the 2012 NSDUH, the extreme weight adjustment step was not necessary either at the DU level or at the person level.

5.1 Phase I Household-Level Weight Components

5.1.1 Weight Components #1 to #7: Selection of a Dwelling Unit

The first seven components in the Phase I sample weights reflect the probability of selecting the DUs. These components were derived from (1) the probability of selecting the census tract within each State SS region, (2) the probability of selecting the segment within each census tract, (3) a quarter segment weight adjustment, (4) a subsegmentation inflation factor, (5) the probability of selecting a DU from within each counted and listed sampled segment, (6) the probability of inclusion of added DUs, and (7) DU percent release adjustment.

Segments were selected with probabilities representing a full year's sample; therefore, Weight Component #3 was set to 1 in the 12-month analysis and was set to 2 in the 6-month analysis (because only half of the segments were used in the analysis). Also, when the field staff, who were responsible for counting and listing, traveled to a specified segment, occasionally they may have found the number of potential DUs to be much greater than what the sample frame (constructed from 2000 U.S. Census Bureau data adjusted for 2005 Claritas projections) indicated. This happened either because of errors in the frame or, more commonly, because of rapid growth in a particular geographic area. When this occurred, the original segment was partitioned and a subsegment was randomly selected. There was an occasional second subsegmentation step when the initial partitioning of segments was insufficient due to out-of-date census counts or the segment was still too large to list after the original subsegmentation. Weight Component #4 (i.e., subsegmentation inflation factor) is an adjustment that accounts for this selection process.

As noted in the 2012 and earlier sample design reports, a lengthy process of determining the optimal DU sample was used during the design of the survey. Weight Component #5 is a result of this process and is equal to the inverse of the DU sample size divided by the total number of DUs counted and listed within a selected segment.

Furthermore, the list of DUs, which includes housing units and group quarters, was constructed by the counting and listing staff during the summer and fall of 2010. Because the listing was done a short time before the 2012 screening and interviewing activities began, no major discrepancies were expected. However, such factors as new construction, demolition, and inaccurate listing were present in some cases. More commonly, DUs may have been "hidden" and, therefore, overlooked by the counter and lister. For all DUs to be given a chance of being selected, the NSDUH has a procedure for locating and adding missed DUs. The current procedure requires field interviewers (FIs) to look both on the property of selected DUs and between each DU and the next listed DU (half-open interval [HOI] rule). In order to make the HOI rule easier to implement in the field, starting from the 2000 survey, the rule was modified such that the HOI would be closed on each map page. Therefore, if the selected DU was the last on a page, the "next listed DU" would be the first one listed on the same page. If the number of added DUs linked to any particular DU did not exceed 5, or if the number for the entire segment was less than or equal to 10, the FI was instructed to consider these DUs as part of his or her assignment. However, if either of these limits was exceeded, the FI would contact RTI for subsampling to be considered. Weight Component #6 accounts for any subsampling that occurred because of added DUs.

To account for corrections, modifications, or both that occurred during the process of design optimization, an additional sample was included throughout all four quarters. Weight Component #7 is the adjustment for the percentage of the DU sample released to FIs in these quarters.

For more detailed information on Weight Components #1 through #7, refer to the 2012 NSDUH sample design report (Morton et al., 2013).

5.1.2 Weight Component #8: Dwelling Unit–Level Nonresponse Adjustment

After DUs were selected, an FI was sent to the DU to screen the residence. Failure to obtain the screening interview from eligible DUs represented the first type of nonresponse encountered in the survey. To account for this nonresponse, as in previous surveys, the (unconditional) sample weights up to this point (equal to the product of Weight Components #1 through #7) were adjusted using a multiplicative adjustment factor derived from modeling response propensity via GEM.

5.1.3 Weight Component #9: Dwelling Unit–Level Poststratification Adjustment

The screener data provided a large sample with information on some demographic variables for the households; therefore, as in two-phase sampling, the screener dwelling unit (SDU) weights first were adjusted for nonresponse and poststratification. Later, estimates for household variables (which were based on screener data) were used as control totals for weight adjustments at the second phase and for person pair-level weights. This was useful because, unlike census controls that were available for individual persons, no controls were available for person pairs. Note that for SDU poststratification, census controls still could be used because each SDU's contribution was computed as the number of persons in the SDU who had certain demographic characteristics multiplied by the SDU weight. It follows that, although explanatory variables used for modeling the weight adjustment were counts instead of binary (0/1), as is often the case, person-level census controls still could be used. For example, age group had five

categories (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older); in SDU poststratification, category 12 to 17 was the number of the persons in this age category within a DU, and so on. The intercept was the total number of persons in the DU, which varied by SDU because SDU size was not constant. Note that when defining interaction control variables for count variables, the corresponding count variables were not simply multiplied, as was done for the binary case; instead, the counts for the category defined by the interaction term (say, Age × Gender) were used.

In addition, the screening process only required the reporting of age for each person rostered; as a result, some fields of demographic information (e.g., race, Hispanic or Latino origin, gender, and two or more races) were missing. Missing data for race and Hispanic or Latino origin were imputed using the predictive mean neighborhood (PMN) methodology (see Appendix C). The probability of observing race (white, black or African American, American Indian or Alaska Native, Asian, and two or more races) was modeled using PROC MULTILOG in SUDAAN®, and the probability of observing Hispanic or Latino origin was modeled using PROC LOGISTIC in SAS. Those probabilities were used in computing predictive means and delta neighborhoods. The "hot deck" method then was used to randomly pick a donor from the neighborhood to impute a missing value for each case. Missing data for gender were imputed using an unweighted hot-deck methodology (see Appendix C). The data file was sorted by auxiliary variables that were considered relevant to the variable being imputed. The sort order of these auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in relation to the variable being imputed. Exhibit 5.3 displays the order in which demographic variables were imputed, along with explanatory variables used in the model or in hot-deck sorting.

Exhibit 5.3 Imputed Demographic Variables and Corresponding Explanatory or Auxiliary Sort Variables

Imputed Variable	Methodology	Explanatory or Auxiliary Sort Variables
Race	Multivariate predictive mean neighborhood (MPMN)	Census region, household type (white, black or African American, Hispanic or Latino), percent of segments that are black or African American, percent of segments that are Hispanic or Latino, percent of owner-occupied dwelling units in segment, segment-combined median rent and housing value, age group
Hispanic or Latino Origin	Univariate predictive mean neighborhood (UPMN)	Census region, imputed race, household type (white, black or African American, Hispanic or Latino), percent of segments that are black or African American, percent of segments that are Hispanic or Latino, percent of owner-occupied dwelling units in segment, segment-combined median rent and housing value, age group
Gender	Hot deck	Census division, imputation-revised Hispanic or Latino origin, imputation-revised race and a random sort number

5.1.4 Weight Component #10: Dwelling Unit–Level Extreme Weight Adjustment

The product of Weight Components #1 through #9 was checked to see if the extreme weight adjustment step was needed. Using the SS region as the domain for the extreme weight definition, weights were defined as extreme if they were outside the range defined by the median

±3 × interquartile range (IQR). Because the unweighted, weighted, and winsorized extreme weight percentages were not high, the extreme weight adjustment was not necessary (see results in Appendix F). Therefore, Weight Component #10 was set to 1 for every DU for which roster information was collected (i.e., every DU with a completed screener).

After this adjustment was completed, the final DU weight was calculated as the product of Weight Components #1 through #10 described previously. This adjusted weight was used to compute household-level estimates from the screener data. It also was used to compute person-level estimates derived from the full roster sample. In addition, these 10 weight components became the first 10 components of the final interview respondent sample weight. The remaining five weight components discussed in the next section account for the person probability of selection for those persons for which a NSDUH interview was sought; they also account for person-level nonresponse, extreme weights, and coverage errors resulting from the last stages of the sample design.

Details on the final models used for DU nonresponse (nr) and poststratification (ps) adjustment for each respective model group can be found in Appendix D.

Table 5.2 presents the weight distribution for design-based weight and unequal weighting effect (UWE) before the implementation of any weight adjustment and after the DU-level nonresponse adjustment and poststratification.

Table 5.2 Weight Distribution for Design-Based Weight and Weight after DU-Level Adjustments

	Minimum	25% Percentile	Median	75% Percentile	Maximum	Mean	*n*	UWE
Design-Based Weight	30	377	520	822	5,182	607	178,586	1.44
Weight after DU-Level Adjustments	9	417	644	1,043	9546	774	153,858	1.54

DU = dwelling unit; UWE = unequal weighting effect.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

5.2 Phase II Person-Level Weight Components

5.2.1 Weight Component #11: Selection of a Person within a Dwelling Unit

The rate at which persons were selected within each DU depended on the age group and was determined during the design of the 2012 study; this also was done for the probabilities of selecting DUs (i.e., Weight Component #5). Note that, similar to the previous surveys, all possible pairs of eligible rostered persons were given some nonzero probability of selection to facilitate unbiased variance estimation. With the use of the Apple Newton handheld computer used by FIs, selection probabilities were adjusted to reflect the total household composition. The survey design restricted the number of interviews to two per DU. With this restriction, a modified Brewer's selection method was used to select either zero, one, or two persons from the DU. (Three ghost units were defined for each DU to allow for the selection of no persons and to avoid division by 0 in Brewer's algorithm.) In short, if the sum of the selection probabilities for all eligible DU members was greater than 2, then the probabilities were ratio-adjusted to sum to 2; sums less than 2 were unadjusted. These adjusted rates then were retained as the final

selection probabilities. An additional design change was made in 2002 and continued through 2012. A new pair-sampling strategy was implemented that increased the number of person pairs selected in DUs with older persons on the roster (Chromy & Penne, 2002). Weight Component #11 represents the inverse of this probability of selection.

5.2.2 Weight Component #12: Selected Person-Level Poststratification Adjustment

The selected person-level poststratification step was started during the 1999 NHSDA. In NHSDAs prior to 1999, a combined step of person-level nonresponse and poststratification to estimated totals from the screener person data was used as a compromise to this step. As was done for the previous surveys, the combined step was divided into two separate steps; the first step was poststratification of the selected persons (i.e., respondents and nonrespondents) to estimated control totals from the screener person data; the second step was respondent person-level nonresponse adjustment (see Component #13) to reproduce control totals from the selected person data (i.e., the full sample). Using two separate steps takes advantage of the inherent two-phase nature of the survey design (although the design is viewed primarily as multistage). With this step, more stable controls for the nonresponse adjustment were obtained (as compared with the traditional nonresponse adjustment) because of the additional selected-person poststratification. Note that this would not have been possible in the absence of screener data on the member demographics of the selected DUs. See Appendix D for details on the final models.

5.2.3 Weight Component #13: Respondent Person-Level Nonresponse Adjustment

The next step was to adjust the sample weights of the interview respondents to the weighted distributions over various demographic domains based on the full sample.

Demographic information for the drug questionnaire respondents was available from two sources—screener data and questionnaire data—while only screener data were available for the large first-phase sample of rostered individuals of all the screened DUs. However, to be consistent with respect to the data source, screener data for both respondents and nonrespondents were used for the person-level nonresponse adjustment. It may be noted that during screening, the only required demographic was the age of each person who was rostered. Thus, such demographics as race/ethnicity and gender of all the rostered eligible persons were not required, and imputation procedures were needed to replace missing data for race/ethnicity and gender. For race/ethnicity, imputations were created using PMN methodology, and for gender, imputations were created using hot-deck methodology. It should be noted that answers from the questionnaire respondents potentially could cause discrepancies between screener values of demographics and their final imputation-revised values. Details on the final models used for the person nonresponse adjustment for each model group can be found in Appendix D.

5.2.4 Weight Component #14: Respondent Person-Level Poststratification Adjustment

This adjustment was to calibrate the weighted respondent-sample data for various demographic domains to the specified control totals obtained from the Census Bureau's estimates of the civilian, noninstitutionalized population aged 12 or older for the year 2012 based on the 2010 census. See Appendix B for details on the derivation of control totals.

After computing the various control totals that were needed, appropriate poststratification factors were applied to the sample weights using GEM to (1) control the resulting UWE and

thereby reduce the potential variance inflation that could result from this weight adjustment, and (2) control for a larger number of main effect and lower order interaction control variables. Details on the final models used for the person-level poststratification adjustment for each model group can be found in Appendix D.

5.2.5 Weight Component #15: Respondent Person-Level Extreme Weight Adjustment

The weights for the product of Weight Components #1 through #14 were checked to see if the extreme weight adjustment step was needed, with extreme weights defined as described in Section 4.1. As in the case of Weight Components #10, unweighted, weighted, and winsorized extreme weight percentages were acceptably low. Therefore, it was decided that the extreme weight adjustment was not required at this stage either. See Appendix G for results. Therefore, Weight Component #15 was set to 1 for each responding person.

Table 5.3 presents the weight distribution and UWE before the implementation of any person-level weight adjustment and after selected person-level poststratification and person-level nonresponse adjustment and poststratification.

Table 5.3 Weight Distribution for Weight before Any Person-Level Adjustment and after Person-Level Adjustments

	Minimum	25% Percentile	Median	75% Percentile	Maximum	Mean	*n*	UWE
Weight before Any Person-Level Adjustment	19	677	1,334	3,483	75,644	2,968	87,656	2.89
Weight after Person-Level Adjustments	1	726	1,549	4,069	133,926	3,807	68,309	3.55

UWE = unequal weighting effect.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

6. Evaluation of Calibration Weights

During the weight calibration process, several criteria for quality control were implemented to assess model adequacy. This chapter describes the individual procedures and presents a summary of their results. All tables referred to in this chapter can be found in Appendices E, F, G, H, and I. More details can be found in the supplement to the appendices.

6.1 Response Rates

Table E.1 in Appendix E displays the final sample sizes for the categories "selected," "eligible," and "completed" at the dwelling unit (DU) level, and for "selected" and "respondents" at the person level from the 2012 National Survey on Drug Use and Health (NSDUH), for both the national and State levels. This table also shows the weighted eligibility rates and weighted response rates for DU screeners and person-level interviews. Table E.1, at the national level, indicates an overall eligibility rate of 83.43 percent as compared with 83.14 percent for 2011. This similarity in overall rates held in nearly all States, with a few notable exceptions: the eligibility rate decreased from 85.38 to 81.98 percent for South Dakota and increased from 68.75 to 73.00 percent for Maine. The screening rate at the national level was also similar for the 2 years (86.07 percent for 2012 vs. 86.98 percent for 2011). The national interview response rate was 72.92 percent, a decrease of 1.39 percentage points compared with 74.31 percent for 2011, with the biggest decrease for Virginia (from 82.23 percent for 2011 to 76.10 percent for 2012) and the biggest increase for Tennessee (from 78.24 percent for 2011 to 80.99 percent for 2012). Table 6.1 presents summary statistics of overall response rates across individual States.

Table 6.1 Summary Statistics of Overall Weighted Response Rates across Individual States

Domain	National Level	Minimum	Median	Maximum
Dwelling Unit Level				
Eligibility Rate	83.43%	73.00% (Maine)	83.62% (Missouri)	88.97% (Virginia)
Screener Response Rate	86.07%	71.89% (New York)	89.57% (Oregon)	94.67% (Utah)
Person Level				
Interview Response Rate	72.92%	63.25% (New York)	74.74% (West Virginia)	83.60% (Utah)

6.2 Percentages of Extreme Weights and Outwinsors

During the stages of modeling adjustments (i.e., nonresponse and poststratification), a major factor in deciding the adequacy of a particular model was the extent of resulting extreme weights among the weights. As explained in Section 4.1, the percentages of extreme weights for the input weight were calculated for some domains of interest prior to adjustment. These values then were compared with the resulting percentages of extreme weights using the product of weight components that included the new adjustment.

Table F.1 in Appendix F and Tables G.1 and G.2 in Appendix G present percentages of extreme weights at both the DU level for the Nation and the person level for the individual States. Unweighted percentages are based on the actual counts of units and are defined as the ratio of extreme weights relative to the total sample size. Weighted percentages reflect the percentage of total extreme value weights relative to the total sample weight, while outwinsor percentages represent the total amount of residual weight (given that the weights are trimmed to the critical values that were used for extreme weight definition) relative to the total sample weight. For evaluation purposes, the outwinsor percentage is considered the most important of the three percentages. This assessment stems from the fact that its value reflects only the actual amount of weight that would be affected if trimming were implemented.

For the 2012 NSDUH sample, domains for extreme weight definitions were defined as follows for various weight adjustments via the generalized exponential model (GEM) (see Section 4.1):

- DU nonresponse by State sampling (SS) region;
- DU poststratification by SS region;
- selected person-level poststratification by SS region and age group,[5] State and age group, SS region, and State;
- person-level nonresponse by SS region and age group, State and age group, SS region, and State; and
- person-level poststratification by SS region and age group, State and age group, SS region, and State.

Before any weight adjustment was implemented, the percentage of unweighted extreme weights was 3.51 percent and the outwinsor was 0.44 percent for the product of design weight components weight 1 to weight 7. After DU-level nonresponse adjustment and poststratification, the percentage of unweighted extreme weights decreased to 1.80 percent and the outwinsor increased to 1.02 percent. When the design weight component weight 11 (inverse probability of selecting a person within a dwelling unit) was introduced, the percentage of unweighted extreme weights increased to 3.41 percent and the outwinsor increased to 1.86 percent. The person-level adjustments, which consisted of selected person-level poststratification, person-level nonresponse adjustment, and person-level poststratification, were able to bring down the percentage of unweighted extreme weights to 1.04 percent and the outwinsor to 0.66 percent.

6.3 Slippage Rates

The slippage rate for a given domain is defined as the percentage difference between the design-based domain population estimate and the census control total, relative to the census control, both before and after poststratification. The tables in Appendix H display national and State-level, domain-specific weight sums for both before and after poststratification. They also present the control totals to be met through poststratification and the relative percentage difference (or the amount of adjustment necessary [positive or negative] to meet the given totals). The first relative difference was used explicitly during the poststratification modeling procedure to identify potential problems for convergence; this was done because large

[5] Age group categories are 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older.

differences in domains with relatively small sample sizes indicate potentially large adjustment factors, which may cause problems in convergence. The reason is that adjustments required for one domain may have an adverse effect for another domain when a unit belongs to both domains.

Consider Table H.11 for Florida, which indicates a sample size of 2,427 for race domain "white"; an initial total, also known as the design-based weight, of 12,427,353; a census total of 13,089,877; and an initial slippage rate of -5.06 percent. The ratio of the census total to the initial total gives the value of the weight adjustment: 1.05. Similar to this example, but in the opposite direction, is Table H.38 for Oklahoma. The domain "Hispanic or Latino" contains a sample size of 133 and an initial slippage rate of 13.75 percent. The initial total of 281,088 and the census total of 247,101 indicate an adjustment of 0.88 would be required.

6.4 Weight Adjustment Summary Statistics

Tables I.1 to I.3 in Appendix I display summary statistics on the product of weight components for before and after all stages of adjustment, for both the DU and person levels. Note that these tables have before and after categories for all adjustments except for the DU poststratification (res.du.ps); this is because the before and after statistics are the same and are, therefore, displayed only as the category after. Note also that there could be changes, although minimal, in person-level specific demographic distributions from screener data to questionnaire data, so the respondent sample unequal weighting effect (UWE) prior to poststratification based on the questionnaire data (e.g., see Table I.1, under the heading "After res.per.nr") would be only slightly different from what would be obtained after the nonresponse adjustment (e.g., see Table I.1, under the heading "Before res.per.ps"). The sample size (n) for the demographic domains from res.per.nr tables also could be different from the res.per.ps tables.

6.5 Sensitivity Analysis of Drug Use Estimates to Baseline Models

In general, there is a trade-off between bias reduction and variance reduction. For instance, with GEM (for nonresponse or poststratification), enlarging a simple model (such as the one with only main effects) has the potential of further reducing the bias. At the same time, this enlargement may be associated with a corresponding increase in the variance of the estimate of the population total. The increased variability comes from estimating the additional parameters included in the model. To check for possible overfitting of the GEM model, a sensitivity analysis was conducted for the poststratification step, where a simple baseline model was fitted with the same bounds and maximum number of iterations as that used for the final, more complex model. Then, point estimates and standard errors (SEs) were examined for substantial changes. If the SE increased only slightly under the complex model or, even better, if it decreased (which is possible because of the correlation between the study and predictor variables), then we would feel comfortable fitting the more complex model.

The SE, a ratio-adjusted estimator denoted by SE1, computed under the DESCRIPT procedure in SUDAAN®, treats the calibration adjustment factors as nonrandom. A more complete method of estimation would take into account the variability present in the weight adjustment. The sandwich formula for the Taylor linearization (see Vaish et al., 2000) is designed to provide an estimate of the variance that adjusts for the random calibration factors to sampling weights via GEM. This "sandwich variance," adjusting for the poststratification variability, is denoted by SE2. Both SE1 and SE2 were calculated, as well as point estimates for

a few important drug recency variables (past year marijuana, alcohol, and cigarette use), across four age groups (12 to17, 18 to 25, 26 to 34, and 35 or older), for the eight States with large sample sizes.

When referring to the standard SUDAAN variance estimator for a survey weighted prevalence estimator, we call it the "naïve Taylor Series" standard error. The sandwich variance, also referred to as the variance estimate from a bias corrected estimating function (BCEF) (Singh & Folsom, 2000), is the "correct" Taylor Series linearization for the survey weighted prevalence estimate when the weights have been calibrated for nonresponse or poststratification. The sandwich variance estimates account for the variance contribution from the weight calibration. It was found in a preliminary study that the naïve Taylor linearization variance is somewhat conservative in comparison with the sandwich variance. The variance estimates of selected outcomes in Tables 6.2 to 6.7 show that, in general, sandwich variances (SE2) are smaller than the naïve Taylor linearization variances (SE1), with a few exceptions. These results confirm the conjecture that BCEF variances, or sandwich variances, are smaller despite the possibility of inflating variance due to adding the weight adjustment variation.

As noted previously, to check for overfitting, the variances of the baseline and final models were compared. In Tables 6.2 to 6.7, there are cases where the SE from the final model is slightly larger than the SE from the baseline model, indicating possible overfitting. However, the variance estimates for the two models (baseline and final) are generally similar to each other. Note that smaller variance estimates for the final model would indicate that the complex model for the poststratification adjustment resulted in better variance reduction (because of correlation between study and predictor variables) and bias reduction (because of meeting control totals corresponding to a number of factor effects). Therefore, the evidence does not favor the view that fitting a large number of parameters in GEM creates instability in estimates.

Table 6.2 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Cigarettes Lifetime											
Total	Point Estimates	62.11	61.93	55.67	55.37	60.94	61.42	61.90	61.64	65.39	65.19
	SE1	0.36	0.36	1.32	1.34	1.40	1.43	1.17	1.17	1.23	1.27
	SE2	0.33	0.32	1.19	1.13	1.31	1.22	1.14	1.04	1.22	1.13
12-17	Point Estimates	17.41	17.40	13.83	13.85	16.24	16.08	16.63	16.53	20.55	20.28
	SE1	0.34	0.35	1.12	1.16	1.22	1.22	1.34	1.29	1.27	1.27
	SE2	0.34	0.34	1.12	1.15	1.22	1.21	1.33	1.29	1.28	1.23
18-25	Point Estimates	59.42	59.53	54.64	55.19	54.31	54.17	62.05	62.71	59.78	59.56
	SE1	0.48	0.49	1.88	1.87	1.65	1.66	1.94	1.88	1.84	1.86
	SE2	0.48	0.47	1.87	1.78	1.64	1.65	1.92	1.74	1.85	1.84
26-34	Point Estimates	69.27	69.00	63.05	63.38	63.06	62.66	69.06	68.73	72.69	72.67
	SE1	0.75	0.78	2.65	2.67	3.29	3.34	2.55	2.66	2.66	2.67
	SE2	0.74	0.74	2.59	2.43	3.27	3.16	2.56	2.41	2.66	2.51
35+	Point Estimates	67.88	67.63	61.03	60.23	67.43	68.22	67.35	66.85	71.88	71.66
	SE1	0.50	0.51	1.99	2.04	1.80	1.83	1.68	1.68	1.77	1.81
	SE2	0.48	0.46	1.89	1.79	1.73	1.69	1.66	1.54	1.76	1.66
Alcohol Lifetime											
Total	Point Estimates	82.50	82.29	79.78	79.62	82.89	83.05	82.78	82.64	85.38	85.25
	SE1	0.26	0.27	0.97	1.00	0.99	1.03	0.94	1.01	0.75	0.77
	SE2	0.23	0.22	0.89	0.85	0.91	0.84	0.89	0.81	0.75	0.67
12-17	Point Estimates	32.41	32.36	30.90	31.18	33.08	33.15	30.65	30.82	32.31	32.17
	SE1	0.44	0.44	1.39	1.43	1.62	1.65	1.59	1.60	1.46	1.46
	SE2	0.43	0.43	1.36	1.37	1.61	1.65	1.57	1.54	1.46	1.43
18-25	Point Estimates	84.36	84.40	84.04	84.28	82.73	82.68	84.39	84.87	85.04	84.98
	SE1	0.36	0.37	1.24	1.25	1.25	1.26	1.35	1.32	1.24	1.24
	SE2	0.36	0.35	1.23	1.19	1.23	1.26	1.33	1.25	1.25	1.20
26-34	Point Estimates	90.67	90.66	87.39	87.67	86.93	86.04	92.64	92.57	93.99	93.87
	SE1	0.50	0.52	1.80	1.82	1.90	2.07	1.44	1.51	1.56	1.63
	SE2	0.49	0.48	1.78	1.75	1.91	2.06	1.44	1.42	1.55	1.47
35+	Point Estimates	87.90	87.57	85.02	84.48	88.49	88.85	88.36	88.00	91.77	91.64
	SE1	0.36	0.38	1.40	1.48	1.26	1.32	1.28	1.42	0.98	1.00
	SE2	0.34	0.32	1.35	1.30	1.21	1.21	1.25	1.16	0.97	0.95

(continued)

Table 6.2 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH (continued)

		New York		Ohio		Pennsylvania		Texas	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Cigarettes Lifetime									
Total	Point Estimates	60.12	59.54	68.41	68.06	66.54	66.20	56.66	56.50
	SE1	1.51	1.59	1.18	1.21	1.23	1.23	1.35	1.36
	SE2	1.42	1.36	1.17	1.19	1.23	1.17	1.23	1.21
12-17	Point Estimates	14.84	15.02	19.31	19.24	18.50	18.68	15.96	15.76
	SE1	1.19	1.23	1.34	1.34	1.39	1.39	1.21	1.22
	SE2	1.19	1.22	1.33	1.31	1.39	1.41	1.22	1.21
18-25	Point Estimates	56.31	57.27	65.18	65.00	62.46	62.35	62.41	61.87
	SE1	1.59	1.67	1.67	1.69	1.59	1.56	1.70	1.78
	SE2	1.58	1.60	1.66	1.61	1.59	1.54	1.71	1.76
26-34	Point Estimates	69.35	66.93	78.00	77.42	78.79	77.54	67.39	66.85
	SE1	3.17	3.40	2.38	2.44	2.47	2.47	2.70	2.82
	SE2	3.15	3.24	2.39	2.36	2.48	2.53	2.65	2.74
35+	Point Estimates	65.11	64.59	74.35	74.05	71.38	71.22	60.00	60.00
	SE1	2.03	2.15	1.65	1.68	1.63	1.66	1.98	2.00
	SE2	1.96	1.88	1.66	1.69	1.64	1.64	1.87	1.84
Alcohol Lifetime									
Total	Point Estimates	83.03	82.17	85.89	85.66	86.57	86.22	77.41	77.09
	SE1	1.12	1.27	0.81	0.83	0.76	0.78	1.11	1.17
	SE2	1.09	1.05	0.79	0.71	0.76	0.73	1.00	0.97
12-17	Point Estimates	34.56	34.81	32.71	32.66	33.31	33.30	30.23	29.62
	SE1	1.55	1.57	1.47	1.48	1.54	1.53	1.50	1.52
	SE2	1.56	1.60	1.46	1.47	1.55	1.53	1.48	1.45
18-25	Point Estimates	86.57	86.32	89.34	89.20	85.97	85.84	83.70	83.34
	SE1	1.46	1.47	0.98	0.98	1.23	1.22	1.35	1.44
	SE2	1.45	1.41	0.97	0.96	1.24	1.19	1.33	1.41
26-34	Point Estimates	92.30	91.67	93.91	94.00	95.97	95.00	90.92	90.62
	SE1	1.63	1.81	1.40	1.40	1.21	1.60	1.62	1.70
	SE2	1.65	1.82	1.39	1.36	1.22	1.61	1.59	1.75
35+	Point Estimates	86.84	85.74	91.43	91.17	92.00	91.77	81.09	80.77
	SE1	1.68	1.89	1.07	1.11	1.05	1.08	1.56	1.62
	SE2	1.65	1.55	1.06	1.02	1.03	0.99	1.49	1.43

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.3 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Marijuana Lifetime											
Total	Point Estimates	42.98	42.77	42.70	42.27	42.58	42.50	41.92	41.77	49.49	49.14
	SE1	0.38	0.38	1.36	1.41	1.63	1.68	1.31	1.30	1.26	1.28
	SE2	0.33	0.32	1.18	1.20	1.55	1.36	1.27	1.12	1.26	1.16
12-17	Point Estimates	17.07	16.97	18.80	18.75	17.64	17.72	16.22	16.00	19.28	19.29
	SE1	0.33	0.34	1.14	1.18	1.21	1.22	1.32	1.28	1.31	1.32
	SE2	0.33	0.33	1.13	1.16	1.20	1.21	1.30	1.24	1.31	1.29
18-25	Point Estimates	52.32	52.19	52.23	52.08	49.08	48.96	53.40	54.01	54.91	54.87
	SE1	0.51	0.51	2.09	2.13	1.74	1.75	1.87	1.83	1.73	1.73
	SE2	0.51	0.49	2.09	2.05	1.73	1.75	1.88	1.82	1.73	1.72
26-34	Point Estimates	55.36	55.18	53.73	53.69	50.55	50.54	52.64	53.30	65.87	65.88
	SE1	0.85	0.88	2.89	2.96	3.43	3.40	2.91	2.97	3.01	2.97
	SE2	0.84	0.80	2.81	2.77	3.39	3.19	2.91	2.85	3.01	2.93
35+	Point Estimates	42.16	41.92	41.65	40.97	43.07	42.94	41.04	40.53	49.76	49.23
	SE1	0.55	0.55	2.04	2.09	2.23	2.31	1.95	1.94	1.78	1.80
	SE2	0.48	0.45	1.83	1.73	2.11	1.86	1.89	1.65	1.79	1.64
Cocaine Lifetime											
Total	Point Estimates	14.58	14.49	17.13	17.17	15.55	15.31	13.90	13.74	15.13	15.05
	SE1	0.26	0.27	1.04	1.05	1.34	1.35	1.02	1.01	1.01	1.00
	SE2	0.25	0.24	0.97	0.94	1.32	1.24	1.01	0.93	1.01	0.96
12-17	Point Estimates	1.10	1.06	1.43	1.37	0.62	0.60	0.97	0.98	0.81	0.79
	SE1	0.09	0.09	0.38	0.37	0.23	0.23	0.28	0.28	0.28	0.27
	SE2	0.09	0.09	0.38	0.36	0.23	0.23	0.28	0.28	0.28	0.26
18-25	Point Estimates	12.38	12.34	13.28	13.45	12.94	12.92	9.68	9.74	9.81	9.77
	SE1	0.32	0.32	1.31	1.31	1.28	1.29	0.94	0.94	1.23	1.24
	SE2	0.32	0.31	1.31	1.28	1.29	1.28	0.94	0.92	1.23	1.23
26-34	Point Estimates	19.46	19.13	18.49	18.43	18.32	18.12	15.36	15.43	21.53	21.46
	SE1	0.65	0.65	2.25	2.29	2.49	2.46	1.88	1.87	2.61	2.57
	SE2	0.64	0.60	2.23	2.15	2.50	2.25	1.88	1.83	2.61	2.55
35+	Point Estimates	16.00	15.95	20.32	20.33	17.40	17.07	16.47	16.19	17.14	17.04
	SE1	0.39	0.39	1.62	1.63	1.84	1.87	1.55	1.53	1.39	1.38
	SE2	0.36	0.35	1.52	1.43	1.81	1.73	1.53	1.39	1.40	1.33

(continued)

Table 6.3 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Lifetime Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH (continued)

		New York		Ohio		Pennsylvania		Texas	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Marijuana Lifetime									
Total	Point Estimates	44.47	43.83	46.54	46.31	42.86	42.32	35.32	35.57
	SE1	1.44	1.45	1.31	1.30	1.38	1.40	1.29	1.31
	SE2	1.37	1.30	1.30	1.11	1.35	1.11	1.17	1.17
12-17	Point Estimates	15.76	16.10	17.49	17.51	16.49	16.44	15.27	15.05
	SE1	1.31	1.34	1.11	1.11	1.27	1.28	1.19	1.19
	SE2	1.31	1.32	1.10	1.07	1.28	1.25	1.19	1.16
18-25	Point Estimates	54.46	55.38	55.50	55.41	54.31	54.46	49.79	49.00
	SE1	1.73	1.70	1.80	1.82	1.59	1.58	1.61	1.67
	SE2	1.72	1.59	1.78	1.74	1.60	1.60	1.57	1.68
26-34	Point Estimates	61.43	58.87	60.48	60.60	61.89	61.16	48.93	49.14
	SE1	3.67	3.78	3.26	3.23	3.18	3.31	2.91	3.02
	SE2	3.65	3.51	3.25	3.04	3.18	3.15	2.86	3.00
35+	Point Estimates	42.21	41.73	46.18	45.90	40.36	39.74	31.99	32.58
	SE1	2.09	2.10	1.80	1.79	1.92	1.95	1.80	1.85
	SE2	2.03	1.93	1.79	1.55	1.86	1.52	1.73	1.75
Cocaine Lifetime									
Total	Point Estimates	14.59	14.27	13.21	13.28	15.45	15.06	12.98	13.04
	SE1	1.02	1.04	0.95	0.95	1.07	1.06	0.81	0.82
	SE2	0.97	0.90	0.95	0.89	1.06	0.92	0.76	0.77
12-17	Point Estimates	0.75	0.76	1.09	1.09	0.71	0.69	2.26	2.25
	SE1	0.27	0.28	0.30	0.30	0.23	0.22	0.47	0.48
	SE2	0.27	0.27	0.30	0.29	0.23	0.22	0.47	0.47
18-25	Point Estimates	12.20	12.60	10.66	10.65	11.31	11.37	14.26	13.73
	SE1	1.18	1.10	0.97	0.98	1.07	1.08	1.16	1.12
	SE2	1.16	1.09	0.97	0.96	1.07	1.08	1.14	1.11
26-34	Point Estimates	19.72	17.65	21.27	21.60	21.91	20.87	22.75	22.73
	SE1	2.53	2.43	2.46	2.44	2.79	2.66	2.46	2.51
	SE2	2.53	2.34	2.46	2.43	2.80	2.64	2.43	2.45
35+	Point Estimates	15.83	15.74	13.84	13.94	16.97	16.61	12.10	12.32
	SE1	1.47	1.53	1.34	1.34	1.50	1.48	1.16	1.21
	SE2	1.41	1.32	1.33	1.26	1.47	1.30	1.13	1.16

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.4 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Cigarettes Past Year											
Total	Point Estimates	26.11	26.07	20.48	20.82	24.19	24.03	28.43	28.30	30.47	30.28
	SE1	0.33	0.34	1.16	1.19	1.29	1.30	1.29	1.29	1.18	1.16
	SE2	0.31	0.30	1.09	1.09	1.28	1.21	1.28	1.22	1.17	1.09
12-17	Point Estimates	11.88	11.83	8.94	8.92	9.86	9.66	10.60	10.76	14.68	14.36
	SE1	0.30	0.30	0.96	0.98	1.04	1.04	0.92	0.94	1.08	1.07
	SE2	0.29	0.29	0.95	0.96	1.04	1.03	0.92	0.94	1.09	1.04
18-25	Point Estimates	40.99	41.04	34.49	34.74	36.78	36.55	41.70	42.23	43.83	43.75
	SE1	0.48	0.48	1.82	1.81	1.63	1.60	2.09	2.06	1.87	1.87
	SE2	0.47	0.46	1.82	1.75	1.61	1.62	2.08	2.00	1.87	1.85
26-34	Point Estimates	38.65	38.23	28.57	28.54	42.10	41.43	38.57	38.37	45.95	46.02
	SE1	0.80	0.81	2.56	2.63	3.19	3.22	2.83	2.82	3.13	3.07
	SE2	0.79	0.77	2.59	2.54	3.18	3.09	2.83	2.79	3.12	2.99
35+	Point Estimates	22.30	22.34	17.07	17.58	20.35	20.29	26.06	25.77	27.10	26.86
	SE1	0.45	0.46	1.63	1.73	1.61	1.65	1.77	1.77	1.67	1.65
	SE2	0.43	0.42	1.58	1.59	1.59	1.58	1.76	1.69	1.67	1.55
Alcohol Past Year											
Total	Point Estimates	66.96	66.74	65.23	65.29	66.84	66.72	69.30	69.26	69.34	69.05
	SE1	0.36	0.37	1.39	1.43	1.32	1.37	1.24	1.28	1.33	1.33
	SE2	0.34	0.33	1.34	1.34	1.26	1.15	1.20	1.10	1.33	1.32
12-17	Point Estimates	26.35	26.29	23.94	24.05	26.85	26.85	25.30	25.45	26.60	26.44
	SE1	0.42	0.42	1.33	1.36	1.62	1.64	1.44	1.47	1.38	1.38
	SE2	0.41	0.41	1.29	1.30	1.61	1.61	1.44	1.44	1.38	1.34
18-25	Point Estimates	77.47	77.42	77.36	77.64	76.38	76.27	78.37	78.81	77.58	77.60
	SE1	0.42	0.43	1.55	1.58	1.43	1.45	1.46	1.45	1.41	1.41
	SE2	0.41	0.40	1.53	1.49	1.40	1.39	1.45	1.41	1.42	1.40
26-34	Point Estimates	80.67	80.52	75.92	76.30	76.61	75.33	84.60	84.86	83.79	83.83
	SE1	0.66	0.70	2.33	2.40	2.54	2.68	2.13	2.14	2.19	2.15
	SE2	0.65	0.67	2.32	2.36	2.54	2.68	2.13	2.02	2.19	2.01
35+	Point Estimates	67.84	67.54	66.57	66.43	68.35	68.41	70.72	70.48	71.28	70.86
	SE1	0.52	0.54	2.00	2.08	1.83	1.91	1.87	1.94	1.89	1.88
	SE2	0.50	0.48	1.99	2.02	1.76	1.64	1.83	1.66	1.90	1.92

(continued)

Table 6.4 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH (continued)

Variables		New York		Ohio		Pennsylvania		Texas	
		Baseline	Final	Baseline	Final	Baseline	Final	Baseline	Final
Cigarettes Past Year									
Total	Point Estimates	24.50	24.10	31.39	31.24	28.69	28.36	24.29	24.10
	SE1	1.26	1.26	1.41	1.41	1.34	1.33	1.25	1.28
	SE2	1.23	1.24	1.38	1.28	1.31	1.14	1.18	1.18
12-17	Point Estimates	9.91	10.00	13.21	13.19	13.36	13.38	10.87	10.84
	SE1	1.01	1.05	1.07	1.08	1.11	1.10	1.04	1.06
	SE2	1.01	1.03	1.06	1.07	1.11	1.11	1.03	1.03
18-25	Point Estimates	38.47	39.35	47.40	47.22	45.03	45.15	42.77	42.23
	SE1	1.56	1.58	1.74	1.75	1.65	1.62	1.64	1.68
	SE2	1.56	1.57	1.74	1.69	1.65	1.63	1.65	1.65
26-34	Point Estimates	39.69	37.11	49.24	48.57	47.71	47.74	36.65	35.98
	SE1	3.36	3.52	3.11	3.12	3.49	3.50	3.06	3.12
	SE2	3.33	3.42	3.12	3.11	3.50	3.48	2.95	2.88
35+	Point Estimates	19.80	19.75	27.23	27.25	23.74	23.27	19.10	19.09
	SE1	1.64	1.69	1.81	1.81	1.69	1.67	1.63	1.66
	SE2	1.63	1.71	1.79	1.70	1.66	1.49	1.60	1.63
Alcohol Past Year									
Total	Point Estimates	70.54	69.99	68.38	68.45	71.81	71.36	61.70	61.66
	SE1	1.38	1.49	1.26	1.27	1.24	1.26	1.46	1.46
	SE2	1.34	1.31	1.26	1.21	1.22	1.15	1.30	1.26
12-17	Point Estimates	30.51	31.01	26.04	26.02	28.56	28.58	24.91	24.40
	SE1	1.43	1.47	1.32	1.33	1.48	1.49	1.43	1.43
	SE2	1.45	1.47	1.32	1.33	1.49	1.50	1.43	1.38
18-25	Point Estimates	82.57	82.31	82.62	82.53	80.43	80.42	75.47	74.74
	SE1	1.56	1.56	1.30	1.31	1.49	1.47	1.61	1.71
	SE2	1.55	1.46	1.29	1.27	1.49	1.44	1.56	1.56
26-34	Point Estimates	82.96	82.58	80.62	81.09	86.30	85.18	80.65	80.16
	SE1	2.27	2.37	2.64	2.54	2.10	2.28	2.28	2.34
	SE2	2.28	2.36	2.63	2.53	2.10	2.29	2.23	2.31
35+	Point Estimates	70.52	69.88	69.31	69.41	73.06	72.67	60.24	60.54
	SE1	2.10	2.25	1.76	1.78	1.80	1.84	2.08	2.08
	SE2	2.05	1.98	1.77	1.73	1.76	1.67	1.96	1.90

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.5 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Marijuana Past Year											
Total	Point Estimates	12.19	12.12	14.11	13.96	11.16	11.28	11.02	11.14	14.67	14.54
	SE1	0.21	0.21	0.86	0.85	0.82	0.85	0.68	0.68	0.73	0.72
	SE2	0.19	0.19	0.80	0.79	0.80	0.80	0.67	0.63	0.72	0.65
12-17	Point Estimates	13.60	13.49	13.96	13.90	13.28	13.35	12.19	12.01	15.99	16.02
	SE1	0.30	0.30	1.04	1.08	1.17	1.18	1.08	1.05	1.21	1.22
	SE2	0.30	0.30	1.04	1.09	1.16	1.17	1.08	1.05	1.21	1.19
18-25	Point Estimates	31.62	31.48	34.25	34.03	29.78	29.61	30.85	31.20	33.15	32.98
	SE1	0.49	0.49	2.00	2.00	1.56	1.58	1.82	1.80	1.52	1.53
	SE2	0.49	0.47	2.01	1.88	1.55	1.58	1.82	1.79	1.53	1.52
26-34	Point Estimates	18.34	18.17	19.14	19.19	17.41	17.68	15.19	15.54	22.69	22.31
	SE1	0.64	0.64	2.26	2.30	2.36	2.41	2.00	2.06	2.89	2.83
	SE2	0.64	0.61	2.22	2.12	2.33	2.18	1.99	1.99	2.89	2.84
35+	Point Estimates	6.48	6.45	8.12	7.95	6.34	6.50	5.72	5.76	9.11	9.02
	SE1	0.25	0.25	1.10	1.08	0.96	1.00	0.88	0.89	0.92	0.91
	SE2	0.24	0.24	1.08	1.04	0.95	1.00	0.87	0.85	0.91	0.87
Cocaine Past Year											
Total	Point Estimates	1.81	1.80	2.33	2.32	2.03	2.09	1.85	1.85	0.98	0.95
	SE1	0.08	0.09	0.32	0.33	0.36	0.40	0.36	0.36	0.20	0.20
	SE2	0.08	0.08	0.32	0.32	0.36	0.38	0.36	0.34	0.20	0.20
12-17	Point Estimates	0.75	0.72	1.00	0.93	0.41	0.39	0.40	0.41	0.26	0.25
	SE1	0.08	0.08	0.31	0.30	0.19	0.19	0.18	0.18	0.13	0.13
	SE2	0.08	0.08	0.31	0.29	0.19	0.19	0.18	0.18	0.13	0.13
18-25	Point Estimates	4.69	4.63	6.46	6.51	5.88	5.87	3.07	3.06	2.54	2.42
	SE1	0.22	0.22	0.94	0.94	0.93	0.92	0.59	0.59	0.59	0.59
	SE2	0.22	0.21	0.94	0.93	0.93	0.91	0.58	0.57	0.59	0.59
26-34	Point Estimates	3.17	3.14	3.43	3.54	5.10	5.20	2.17	2.17	1.82	1.84
	SE1	0.31	0.31	0.97	1.01	1.42	1.48	0.80	0.79	0.83	0.84
	SE2	0.31	0.30	0.97	1.01	1.42	1.46	0.80	0.78	0.83	0.82
35+	Point Estimates	1.05	1.06	1.30	1.26	0.96	1.04	1.75	1.74	0.60	0.58
	SE1	0.10	0.10	0.42	0.42	0.38	0.45	0.51	0.50	0.26	0.25
	SE2	0.10	0.10	0.41	0.41	0.39	0.43	0.51	0.48	0.26	0.25

(continued)

Table 6.5 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Year Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH (continued)

		New York		Ohio		Pennsylvania		Texas	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Marijuana Past Year									
Total	Point Estimates	14.68	14.81	13.58	13.64	11.15	11.14	9.96	9.89
	SE1	0.99	1.01	0.77	0.77	0.78	0.77	0.65	0.66
	SE2	0.95	1.01	0.76	0.71	0.75	0.64	0.60	0.59
12-17	Point Estimates	13.31	13.56	14.57	14.59	13.96	13.89	12.59	12.35
	SE1	1.19	1.21	1.06	1.06	1.15	1.15	1.09	1.10
	SE2	1.18	1.20	1.05	1.05	1.15	1.12	1.08	1.04
18-25	Point Estimates	36.38	37.18	30.97	30.74	32.93	33.08	26.08	25.71
	SE1	1.64	1.68	1.65	1.65	1.65	1.64	1.34	1.36
	SE2	1.66	1.67	1.64	1.59	1.66	1.68	1.31	1.31
26-34	Point Estimates	23.45	23.24	20.77	21.07	22.14	21.72	17.58	17.47
	SE1	2.90	3.01	2.78	2.71	2.88	2.85	2.19	2.26
	SE2	2.91	3.02	2.78	2.73	2.88	2.80	2.15	2.20
35+	Point Estimates	7.95	8.17	8.50	8.60	4.29	4.32	3.61	3.67
	SE1	1.12	1.19	0.92	0.94	0.75	0.76	0.59	0.62
	SE2	1.11	1.36	0.92	0.89	0.74	0.69	0.59	0.61
Cocaine Past Year									
Total	Point Estimates	2.54	2.50	1.73	1.75	1.91	1.95	1.50	1.46
	SE1	0.35	0.32	0.31	0.32	0.34	0.37	0.18	0.19
	SE2	0.35	0.33	0.31	0.31	0.34	0.35	0.18	0.19
12-17	Point Estimates	0.61	0.62	0.33	0.34	0.29	0.27	1.65	1.63
	SE1	0.25	0.26	0.15	0.15	0.15	0.14	0.41	0.42
	SE2	0.25	0.26	0.15	0.15	0.15	0.14	0.41	0.41
18-25	Point Estimates	5.90	6.39	3.73	3.66	3.62	3.67	4.45	4.16
	SE1	0.76	0.82	0.59	0.60	0.58	0.59	0.58	0.58
	SE2	0.74	0.80	0.59	0.59	0.58	0.60	0.58	0.60
26-34	Point Estimates	4.57	4.17	3.08	2.94	2.97	3.23	2.19	2.22
	SE1	1.43	1.18	1.03	0.97	1.07	1.29	0.77	0.79
	SE2	1.44	1.20	1.03	0.95	1.07	1.25	0.77	0.78
35+	Point Estimates	1.58	1.54	1.26	1.34	1.58	1.59	0.58	0.59
	SE1	0.39	0.39	0.41	0.45	0.45	0.46	0.22	0.22
	SE2	0.40	0.40	0.41	0.42	0.45	0.44	0.21	0.21

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.6 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Cigarettes Past Month											
Total	Point Estimates	22.17	22.12	16.30	16.64	20.33	20.22	24.01	23.89	26.66	26.54
	SE1	0.31	0.32	1.05	1.09	1.25	1.27	1.19	1.19	1.19	1.18
	SE2	0.30	0.29	1.02	1.02	1.25	1.20	1.18	1.13	1.19	1.12
12-17	Point Estimates	6.64	6.55	4.40	4.35	5.45	5.31	5.58	5.68	8.62	8.53
	SE1	0.22	0.22	0.69	0.69	0.83	0.81	0.72	0.74	0.81	0.82
	SE2	0.22	0.21	0.68	0.68	0.84	0.81	0.72	0.74	0.81	0.80
18-25	Point Estimates	31.78	31.82	25.15	25.24	26.82	26.89	32.94	33.27	34.54	34.44
	SE1	0.46	0.46	1.60	1.60	1.45	1.48	1.74	1.73	1.95	1.94
	SE2	0.45	0.45	1.60	1.57	1.44	1.54	1.73	1.69	1.95	1.93
26-34	Point Estimates	32.90	32.60	22.06	21.91	37.69	37.12	33.63	33.45	41.13	40.90
	SE1	0.77	0.78	2.32	2.35	3.10	3.09	2.87	2.84	3.22	3.17
	SE2	0.76	0.74	2.34	2.27	3.10	2.96	2.86	2.81	3.21	3.03
35+	Point Estimates	20.08	20.08	14.75	15.32	17.78	17.69	22.77	22.54	24.96	24.85
	SE1	0.43	0.44	1.52	1.62	1.60	1.64	1.64	1.65	1.63	1.62
	SE2	0.41	0.40	1.49	1.50	1.59	1.58	1.63	1.57	1.62	1.53
Alcohol Past Month											
Total	Point Estimates	52.25	52.10	49.18	49.05	51.64	51.83	55.16	55.15	54.94	54.38
	SE1	0.39	0.40	1.42	1.47	1.53	1.56	1.35	1.38	1.41	1.41
	SE2	0.37	0.35	1.38	1.35	1.46	1.35	1.31	1.22	1.41	1.35
12-17	Point Estimates	12.96	12.90	11.54	11.33	12.86	12.97	12.20	12.31	13.96	13.87
	SE1	0.32	0.32	1.02	1.03	1.14	1.17	1.16	1.18	0.99	1.01
	SE2	0.31	0.31	1.01	1.00	1.14	1.12	1.16	1.17	0.99	0.97
18-25	Point Estimates	60.28	60.21	58.53	58.62	58.53	58.72	63.88	64.55	62.30	62.31
	SE1	0.49	0.50	1.79	1.82	1.75	1.76	1.72	1.71	1.72	1.72
	SE2	0.49	0.48	1.77	1.71	1.72	1.66	1.71	1.68	1.73	1.73
26-34	Point Estimates	64.65	64.46	59.99	60.19	58.82	57.74	67.39	66.87	68.72	68.36
	SE1	0.81	0.84	2.83	2.91	3.10	3.16	2.93	2.97	2.93	2.93
	SE2	0.81	0.81	2.81	2.82	3.09	3.07	2.93	2.74	2.94	2.85
35+	Point Estimates	53.74	53.58	50.52	50.22	53.96	54.38	57.23	57.17	56.93	56.15
	SE1	0.56	0.57	1.99	2.06	2.09	2.14	2.00	2.04	1.98	1.96
	SE2	0.53	0.50	1.97	1.95	2.02	1.89	1.96	1.79	1.98	1.90

(continued)

Table 6.6 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Licit Drug Estimates, Cigarettes and Alcohol: 2012 NSDUH (continued)

Variables		New York		Ohio		Pennsylvania		Texas	
		Baseline	Final	Baseline	Final	Baseline	Final	Baseline	Final
Cigarettes Past Month									
Total	Point Estimates	20.91	20.57	27.86	27.74	25.46	25.17	19.77	19.60
	SE1	1.20	1.20	1.32	1.32	1.29	1.28	1.10	1.12
	SE2	1.17	1.17	1.30	1.22	1.27	1.11	1.06	1.07
12-17	Point Estimates	4.99	5.10	8.56	8.57	7.89	7.80	5.63	5.54
	SE1	0.71	0.75	0.81	0.81	0.83	0.82	0.79	0.79
	SE2	0.71	0.74	0.81	0.82	0.83	0.81	0.78	0.76
18-25	Point Estimates	30.30	31.16	38.65	38.53	36.38	36.48	31.85	31.22
	SE1	1.66	1.61	1.77	1.77	1.52	1.49	1.71	1.71
	SE2	1.65	1.60	1.76	1.73	1.52	1.51	1.73	1.73
26-34	Point Estimates	32.89	31.10	44.44	43.86	42.30	42.67	31.66	30.95
	SE1	2.95	3.13	3.04	3.02	3.46	3.46	2.88	2.93
	SE2	2.92	3.09	3.04	3.01	3.47	3.46	2.78	2.69
35+	Point Estimates	18.18	18.00	25.16	25.19	22.32	21.86	16.38	16.43
	SE1	1.58	1.60	1.71	1.71	1.65	1.63	1.45	1.47
	SE2	1.57	1.62	1.69	1.61	1.62	1.46	1.43	1.46
Alcohol Past Month									
Total	Point Estimates	56.30	55.91	53.22	53.22	56.25	55.61	46.76	46.65
	SE1	1.52	1.57	1.41	1.41	1.50	1.54	1.39	1.38
	SE2	1.48	1.39	1.40	1.35	1.47	1.41	1.30	1.26
12-17	Point Estimates	16.25	16.56	13.08	13.08	13.59	13.59	12.22	11.95
	SE1	1.19	1.23	1.00	1.02	1.10	1.10	1.00	1.01
	SE2	1.19	1.20	1.00	1.01	1.11	1.11	1.01	0.99
18-25	Point Estimates	65.06	65.64	64.40	64.31	64.68	64.90	57.14	56.25
	SE1	1.74	1.74	1.64	1.65	1.94	1.93	1.93	1.95
	SE2	1.76	1.68	1.62	1.57	1.95	1.88	1.90	1.81
26-34	Point Estimates	71.58	69.74	66.63	66.76	70.07	68.95	60.71	60.23
	SE1	2.72	2.94	2.85	2.84	3.31	3.46	2.90	2.93
	SE2	2.75	2.85	2.84	2.78	3.30	3.28	2.91	2.95
35+	Point Estimates	56.31	56.13	54.18	54.25	57.60	56.87	47.00	47.17
	SE1	2.25	2.33	1.92	1.92	2.08	2.16	2.04	2.01
	SE2	2.19	2.04	1.93	1.87	2.04	2.00	1.95	1.86

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.7 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH

		United States		California		Florida		Illinois		Michigan	
Variables		**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**	**Baseline**	**Final**
Marijuana Past Month											
Total	Point Estimates	7.31	7.25	9.37	9.23	6.77	6.94	6.26	6.37	9.25	9.12
	SE1	0.17	0.16	0.74	0.73	0.68	0.70	0.51	0.52	0.59	0.58
	SE2	0.16	0.15	0.72	0.70	0.67	0.68	0.50	0.47	0.59	0.56
12-17	Point Estimates	7.30	7.22	7.08	7.02	6.85	7.07	4.97	5.11	9.82	9.92
	SE1	0.23	0.22	0.82	0.83	0.85	0.89	0.65	0.68	0.93	0.95
	SE2	0.23	0.22	0.81	0.82	0.85	0.88	0.65	0.68	0.93	0.93
18-25	Point Estimates	18.86	18.75	22.61	22.35	18.38	18.43	19.95	20.23	22.24	22.00
	SE1	0.39	0.39	1.63	1.63	1.29	1.32	1.49	1.49	1.54	1.53
	SE2	0.38	0.38	1.64	1.54	1.29	1.32	1.51	1.52	1.54	1.52
26-34	Point Estimates	11.48	11.30	11.92	11.90	12.13	12.49	10.31	10.46	14.58	14.32
	SE1	0.52	0.51	1.83	1.84	1.99	2.06	1.75	1.77	2.37	2.29
	SE2	0.52	0.49	1.79	1.75	1.99	1.94	1.75	1.74	2.36	2.29
35+	Point Estimates	3.93	3.91	5.99	5.85	3.64	3.81	2.65	2.70	5.45	5.36
	SE1	0.21	0.21	1.02	0.99	0.80	0.84	0.57	0.59	0.73	0.72
	SE2	0.20	0.19	1.01	0.96	0.80	0.83	0.57	0.56	0.73	0.71
Cocaine Past Month											
Total	Point Estimates	0.63	0.63	0.73	0.73	0.68	0.72	0.87	0.85	0.42	0.41
	SE1	0.05	0.05	0.20	0.20	0.24	0.28	0.24	0.23	0.16	0.16
	SE2	0.05	0.05	0.20	0.20	0.24	0.27	0.24	0.23	0.16	0.16
12-17	Point Estimates	0.13	0.12	0.23	0.15	0.13	0.14	0.00	0.00	0.00	0.00
	SE1	0.03	0.03	0.14	0.09	0.09	0.10	0.00	0.00	0.00	0.00
	SE2	0.03	0.03	0.14	0.09	0.09	0.10	0.00	0.00	0.00	0.00
18-25	Point Estimates	1.19	1.15	1.23	1.25	1.80	1.73	1.07	1.02	0.81	0.82
	SE1	0.10	0.10	0.38	0.38	0.45	0.42	0.32	0.32	0.38	0.39
	SE2	0.10	0.10	0.38	0.38	0.45	0.42	0.32	0.31	0.38	0.39
26-34	Point Estimates	1.19	1.15	1.02	1.09	0.97	0.92	1.43	1.38	0.26	0.26
	SE1	0.21	0.20	0.52	0.56	0.57	0.55	0.69	0.67	0.26	0.26
	SE2	0.21	0.19	0.52	0.55	0.57	0.55	0.69	0.67	0.26	0.26
35+	Point Estimates	0.47	0.49	0.63	0.61	0.49	0.57	0.83	0.81	0.43	0.42
	SE1	0.07	0.07	0.29	0.29	0.33	0.40	0.33	0.33	0.24	0.23
	SE2	0.07	0.07	0.29	0.29	0.33	0.38	0.33	0.32	0.24	0.23

(continued)

Table 6.7 Point Estimates, Ratio-Adjusted Standard Errors (SE1), and Sandwich Standard Errors (SE2) for Baseline and Final Models—Drug Estimates (United States and Eight Large States): Past Month Illicit Drug Estimates, Marijuana and Cocaine: 2012 NSDUH (continued)

Variables		New York		Ohio		Pennsylvania		Texas	
		Baseline	Final	Baseline	Final	Baseline	Final	Baseline	Final
Marijuana Past Month									
Total	Point Estimates	8.88	8.82	8.05	8.16	5.88	5.92	5.15	5.13
	SE1	0.69	0.72	0.55	0.56	0.48	0.49	0.48	0.49
	SE2	0.66	0.64	0.55	0.52	0.47	0.42	0.46	0.47
12-17	Point Estimates	8.03	8.06	8.36	8.34	7.23	7.22	6.74	6.55
	SE1	1.01	1.02	0.78	0.78	0.86	0.86	0.82	0.82
	SE2	1.01	1.01	0.78	0.78	0.86	0.85	0.82	0.78
18-25	Point Estimates	22.24	22.59	19.18	19.13	18.08	18.22	13.70	13.45
	SE1	1.37	1.39	1.45	1.45	1.19	1.21	0.97	0.98
	SE2	1.38	1.40	1.45	1.40	1.19	1.23	0.99	0.99
26-34	Point Estimates	14.93	13.86	12.89	13.42	12.79	12.87	8.99	8.93
	SE1	2.39	2.24	2.02	2.01	2.06	2.12	1.70	1.74
	SE2	2.39	2.16	2.02	2.01	2.06	2.08	1.70	1.77
35+	Point Estimates	4.58	4.77	4.80	4.89	1.92	1.93	1.80	1.89
	SE1	0.83	0.93	0.68	0.71	0.45	0.44	0.45	0.48
	SE2	0.82	0.86	0.68	0.68	0.44	0.42	0.46	0.48
Cocaine Past Month									
Total	Point Estimates	0.80	0.75	0.56	0.56	0.74	0.71	0.58	0.58
	SE1	0.21	0.19	0.19	0.19	0.23	0.23	0.12	0.13
	SE2	0.21	0.19	0.19	0.19	0.23	0.22	0.12	0.13
12-17	Point Estimates	0.31	0.32	0.13	0.12	0.07	0.07	0.34	0.34
	SE1	0.18	0.18	0.09	0.09	0.07	0.07	0.16	0.16
	SE2	0.18	0.18	0.09	0.09	0.07	0.07	0.16	0.15
18-25	Point Estimates	1.10	1.18	1.08	1.04	0.81	0.81	1.15	1.03
	SE1	0.31	0.33	0.39	0.38	0.30	0.30	0.26	0.24
	SE2	0.32	0.34	0.39	0.37	0.30	0.30	0.26	0.29
26-34	Point Estimates	1.50	1.23	0.86	0.81	2.14	1.95	1.37	1.40
	SE1	0.97	0.82	0.51	0.48	0.90	0.82	0.64	0.66
	SE2	0.97	0.80	0.50	0.47	0.90	0.82	0.64	0.66
35+	Point Estimates	0.64	0.61	0.46	0.48	0.53	0.54	0.29	0.29
	SE1	0.24	0.23	0.27	0.28	0.29	0.30	0.14	0.15
	SE2	0.24	0.23	0.27	0.27	0.29	0.29	0.14	0.14

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

References

Chen, P., Penne, M. A., & Singh, A. C. (2000). Experience with generalized exponential model (GEM) for weight calibration for NHSDA. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 604-607). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Chromy, J. R., & Penne, M. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY [CD-ROM]* (pp. 552-554). Alexandria, VA: American Statistical Association.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376-382.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Folsom, R. E., & Witt, M. B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. In *Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Toronto, Ontario, Canada* (pp. 428-433). Alexandria, VA: American Statistical Association.

Frechtel, P., Carpenter, L., Edwards, S., Gulledge, K., Handley, W., King, S., Kroutil, L., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2014). Editing and imputation report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.107.006.007). Research Triangle Park, NC: RTI International.

Kish, L. (1965). *Survey sampling*. New York: John Wiley & Sons.

Little, R. J. A., & Rubin, D. B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

Singh, A. C., & Folsom, R. E., Jr. (2000). Bias corrected estimating function approach for variance estimation adjusted for poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 600-615). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Singh, A. C., Penne, M. A., & Gordek, H. (1999, September 27). *Poststratification and nonresponse adjustments for the six month analysis of the 1999 NHSDA data: A discussion paper* [memo to SAMHSA]. Research Triangle Park, NC: RTI.

Singh, A., Grau, E., & Folsom, R., Jr. (2002). Predictive mean neighborhood imputation for NHSDA substance use data. In J. Gfroerer, J. Eyerman, & J. Chromy (Eds.), *Redesigning an ongoing national household survey: Methodological issues* (pp. 111-133, DHHS Publication No. SMA 03-3768). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Vaish, A. K., Gordek, H., & Singh, A. C. (2000). Variance estimation adjusted for weight calibration via the generalized exponential model with application to the National Household Survey on Drug Abuse. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 616-621). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/srms/proceedings/]

Appendix A: Technical Details about the Generalized Exponential Model

Appendix A: Technical Details about the Generalized Exponential Model

A.1 Distance Function

Let $\Delta(w,d)$ denote the distance between the initial weights $d = \{d_k : k \in s\}$ and the adjusted weights *w*, with *k* being the kth unit in the sample and *s* being the sample selected. The distance function minimized under the generalized exponential model (GEM), subject to calibration constraints, is given by

$$\Delta(w,d) = \sum_{k\in s} \frac{d_k}{A_k} \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\}, \tag{A.1.1}$$

where $a_k = w_k / d_k, A_k = (u_k - \ell_k)/[(u_k - c_k)(c_k - \ell_k)]$ and ℓ_k, c_k, and u_k are prescribed real numbers. Let T_x denote the *p*-vector of control totals corresponding to predictor variables ($x_1, ..., x_p$). Then, the calibration constraints for the above minimization problem are

$$\sum_{k\in s} x_k d_k a_k = T_x. \tag{A.1.2}$$

The solution for the above minimization problem, if it exists, is given by a GEM with model parameters λ; that is,

$$a_k(\lambda) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}{(u_k - c_k) + (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}. \tag{A.1.3}$$

Note that the number of parameters in the GEM should be $\leq n$, where *n* is the size of the sample *s*. This is also the dimension of vectors *d* and *w*. It follows from Equation A.1.3 that

$$\ell_k < a_k < u_k, k = 1, \ldots, n. \tag{A.1.4}$$

The usual raking ratio method (Singh & Mohl, 1996) of weight adjustment is a special case of the GEM, noting that for $\ell_k = 0$, $u_k = \infty$, $c_k = 1$, and $k = 1, \ldots, n$, we have

$$\Delta(w,d) = \sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k (a_k - 1) \tag{A.1.5}$$

and $a_k(\lambda) = \exp(x_k' \lambda)$.

The logit method of Deville and Särndal (1992) is also a special case of the GEM, by setting $\ell_k = \ell$, $u_k = u$, and $c_k = 1$ for all *k*. The new method was introduced by Folsom and Singh (2000).

A.2 GEM Adjustments for Extreme Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (A.1.3) can be justified for all three types of adjustment: extreme value treatment, nonresponse, and poststratification. For extreme value treatment via winsorization, denote the winsorized weights by $\{b_k\}$, where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = \text{med}\{d_k\} \pm 3 * \text{IQR}$ if d_k is an extreme weight, where IQR denotes the interquartile range, and the median and quartiles for the weights are defined with respect to a suitable design-based stratum.

For the nonresponse adjustment, the sample is first divided into two parts: the nonextreme weight subsample and the extreme weight subsample. For nonextreme weights, the following are set: $\ell_2 = 1$, $c_2 = \rho^{-1}$, $u_2 = u > \rho^{-1}$, where ρ is the overall response propensity. For extreme weights with high weights, $\ell_k = \ell_1 m_k$, $c_k = \rho^{-1} m_k$, and $u_k = u_1 m_k$, where $m_k = b_k / d_k$ and $1 \le \ell_1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, for extreme weights with low weights, $\ell_k = \ell_3 m_k$, $c_k = \rho^{-1} m_k$, $u_k = u_3 m_k$, and $1 \le \ell_3 < \rho^{-1} = c_3 < u_3$.

For the poststratification adjustment, the following weights are set: for nonextreme weights, $\ell_k = \ell_2$, $c_k = c_2 = 1$, and $u_k = u_2$; for high extreme weights, $\ell_k = \ell_1 m_k$, $c_k = m_k$, and $u_k = u_1 m_k$; and similarly, for low extreme weights, $\ell_k = \ell_3 m_k$, $c_k = m_k$, and $u_k = u_3 m_k$. The extreme value adjustment is identical to poststratifcation, except for tighter bounds on extreme weights resulting from the final poststratification.

Notice that the GEM allows the flexibility of specifying different bounds for different subsamples. In addition, the lower bound (in the case of nonresponse adjustments) can be made to equal one by choosing the center $c_k > 1$.

A.3 Newton-Raphson Steps

Let X denote the $n \times p$ matrix of predictor values, and for the vth iteration,

$$\Gamma_{\phi v} = \text{diag}\left(d_k \phi_k^{(v)}\right), \phi_k^{(o)} = 1,$$

where $\phi_k^{(v)} = \left[\left(u_k - a_k^{(v)}\right)\left(a_k^{(v)} - \ell_k\right)\right] \Big/ \left[\left(u_k - c_k\right)\left(c_k - \ell_k\right)\right]$.

Then, for the Newton-Raphson iteration v, the value of the p-vector λ is adjusted as

$$\lambda^{(v)} = \lambda^{(v-1)} + \left(X'\Gamma_{\phi,v-1}X\right)^{-1}\left(T_x - \hat{T}_x^{(v-1)}\right),$$

where $\lambda^{(0)} = 1$.

The convergence criterion is based on the Euclidean distance $\left\|T_x - \hat{T}_x^{(v)}\right\|$, which is defined as $\sqrt{\left(T_x - \hat{T}_x^{(v)}\right)'\left(T_x - \hat{T}_x^{(v)}\right)}$. At each iteration, it is checked to determine whether it is decreasing. If it is not, a half step is used in the iteration increment.

A.4 Scaled Constrained Exponential Model

In National Household Surveys on Drug Abuse (NHSDAs)[1] prior to 1999, constrained exponential models (CEMs) were used for poststratification, and scaled CEMs were used for nonresponse adjustments. The CEM refers to the logit model of Deville and Särndal (1992), in which lower and upper bounds do not vary with k; that is, $\ell_k = \ell$, $u_k = u$, and $c_k = c = 1$, such that $\ell < 1 < u$. Thus, the CEM is a special case of the GEM. For the nonresponse adjustment, Folsom and Witt (1994) modified the CEM estimating equations by a scaling factor (ρ^{-1}, the inverse of the overall response propensity), such that $1 < \rho^{-1}a_k < \rho^{-1}u$. This implies that choosing ℓ in the CEM as ρ ensures that the scaled adjustment factor for nonresponse is at least one.

[1] The National Household Survey on Drug Abuse (NHSDA) was renamed the National Survey on Drug Use and Health (NSDUH) in the 2002 survey year.

Appendix B: Poststratification Control Totals

Appendix B: Poststratification Control Totals

For poststratification, quarterly State-specific totals for the target population (civilian, noninstitutionalized, aged 12 or older) are required for 120 demographic domains defined by Age, Race, Gender, and Hispanicity (6 × 5 × 2 × 2) (Exhibit B.1). The Population Estimates Branch of the U.S. Bureau of the Census produced, in response to a special request, the necessary population estimates based on monthly State-level estimates of the target population, which were based on the enumerated population from Census 2010.

To arrive at quarterly estimates, approximations at the midpoints of the quarters were needed. To get these approximations, the estimates from the last 2 months in each quarter were averaged. For example, to obtain an approximation for the first quarter of 2012, the U.S. census estimates for February 1 and March 1 were averaged, resulting in a population estimate appropriate for February 15 (i.e., the midpoint of Quarter 1).

Exhibit B.1 Definition of Levels for Variables

Age (years)
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50-64, 6: 65+

Race
1: White, 2: Black or African American, 3: American Indian or Alaska Native, 4: Asian or Native Hawaiian or Pacific Islander, 5: Two or More Races

Gender
1: Male, 2: Female

Hispanicity
1: Hispanic or Latino, 2: Non-Hispanic or Latino

Appendix C: Imputation Methodology

Appendix C: Imputation Methodology

C.1 Unweighted Hot Deck

The adjustments of (1) dwelling unit (DU) poststratification, (2) poststratification of the selected sample to all eligible rostered persons, and (3) person-level nonresponse required the use of demographic information obtained from the 2012 National Survey on Drug Use and Health (NSDUH) screener interview. However, at the time of screening, the only required information for an individual was age, and, thus, some demographic information (i.e., Gender, Hispanic or Latino origin, and race) was missing. Therefore, some form of imputation was required for cases with missing data.[1] This imputation was performed using an unweighted hot-deck methodology. The unweighted hot-deck method of imputing a variable with missing responses (which is called the base variable in this appendix) involved three basic steps.

1. *Forming imputation classes.* When a strong logical association existed between the base variable and certain auxiliary variables, the data set was partitioned by the auxiliary variables, and imputation procedures were implemented independently within classes defined by the cross of the auxiliary variables.
2. *Sorting the file.* Within each imputation class, the file was sorted by auxiliary variables that were relevant to the item being imputed. The sort order of the auxiliary variables was chosen to reflect the degree of importance of the auxiliary variables in relation to the base variable being imputed (i.e., those auxiliary variables that were better predictors for the item being imputed were used as the first sorting variables).

   For the 2012 NSDUH, two types of sorting procedures were used to sort the files prior to imputation:

   (a) Straight Sort. A set of variables was sorted in ascending order by the first variable specified, then, within each level of the first variable, the file was sorted in ascending order by the second variable specified, and so on. For example:

1	1	1
1	1	2
1	2	1
1	2	2
1	3	1
1	3	2
2	1	1
2	1	2
2	2	1
2	2	2
2	3	1
2	3	2

[1] Because the imputation of these demographic variables was not required for the main NSDUH analysis, it is documented here in the weighting report.

(b) Serpentine Sort. A set of variables was sorted so that the direction of the sort (ascending or descending) changed each time the value of a variable changed. For example:

1	1	1
1	1	2
1	2	2
1	2	1
1	3	1
1	3	2
2	3	2
2	3	1
2	2	1
2	2	2
2	1	2
2	1	1

The serpentine sort has the advantage of minimizing the change in the entire set of auxiliary variables whenever any one of the variables changes its value.

3. *Replace missing values.* The file was sorted and then read sequentially. Each time an item respondent was encountered (i.e., the base variable was nonmissing), the base variable response was stored, updating the donor response, and any subsequent nonrespondent encountered received the stored donor response, creating the statistically imputed response. A starting value was needed if an item nonrespondent was the first record on a sorted file. Typically, the response from the first respondent on the sorted file was used as the starting value.

   Note that because the file was sorted by relevant auxiliary variables, the preceding item respondent (donor) closely matched the neighboring item nonrespondent (recipient) with respect to the auxiliary variables.

For more information on the general hot-deck method of item imputation, see Little and Rubin, 1987 (pp. 62-67).

With the unweighted sequential hot-deck imputation procedure, for any particular item being imputed, there was the risk of several nonrespondents appearing next to one another on the sorted file. To detect this problem in NSDUH, for every variable being imputed, a record was kept of the imputation donor. Then, by examining frequencies by imputation donor, if several nonrespondents were lining up next to one another in the sort, the situation could be detected. When this problem occurred, sort variables were added or eliminated, or the order of the sort variables was rearranged.

C.2 Predictive Mean Neighborhood (PMN)

As in 2002, the predictive mean neighborhood (PMN) methodology was used for the 2012 NSDUH weighting process to impute "race" and "Hispanic or Latino origin" for the screener demographic information, as well as the questionnaire data (Singh, Grau, & Folsom, 2002). Because there was not a good set of predictors for PMN modeling, the unweighted

sequential hot-deck method was used to impute gender. Unweighted sequential hot deck is simple and quick to implement, but it has a number of disadvantages:

- The first few sorting covariates almost entirely determine what donor will be used for a particular respondent with missing data, regardless of how many sorting covariates are included.
- There is no mechanism derived from the data to weight the sorting covariates based on their relationship to the response variable.
- Weights are not used to determine the most appropriate donor for a respondent with missing data.
- The correlations across multiple outcome variables imputed to the same record are not accounted for when finding a donor.
- The choice of donor, after the sort has been completed, may be deterministic; this may introduce bias in estimating means and totals and, thus, make it difficult to determine the variance of the estimator when taking imputation into account.

To address the deficiencies of the unweighted sequential hot deck, the PMN methodology was developed for NSDUH. It is a combination of two commonly used imputation methods: a nonmodel-based hot deck and the model-based predictive mean matching method of Rubin. It enhances the predictive mean matching method in that it can be applied to both discrete and continuous variables either individually or jointly. It also enhances the nearest neighbor hot-deck method in that the distance function used to find neighbors is no longer ad hoc. It is easily applicable to problems of both univariate (UPMN) and multivariate (MPMN) imputations. Univariate imputation is used for imputing a single continuous or dichotomous discrete variable independently, while multivariate imputation arises when values of two or more variables are missing for a single respondent or when a single polytomous variable has missing values. (A polytomous variable is a categorical variable with three or more possible values, such as marital status, which is categorical and has the possible values of married, widowed, divorced, and never married.)

The procedure for implementing univariate and multivariable imputations can be summarized with the following six steps. Steps 2 through 5, and sometimes Step 6, were cycled through each of the variables in the order determined by Step 1. Steps 4 and 5 (Steps 4 through 6, when applicable) could be considered a variant of a random nearest neighbor hot deck.

Step 1: Hierarchy definition. Determine the order in which variables are modeled, so that variables early in the hierarchy may be used for modeling the conditional predictive mean (i.e., variables early in the hierarchy have the potential to be part of the set of covariates for variables later in the hierarchy).

For each variable:

Step 2: Setup for model building and hot-deck assignment. For each model that is fitted, two groups must be created: complete and incomplete data respondents (item respondents and item

nonrespondents). Complete data respondents have complete data across the variables of interest, and incomplete data respondents encompass the remainder of respondents.

Step 3: Sequential hierarchical modeling. The model is built using the complete data for respondents only, with weights adjusted for item nonresponse.

Step 4: Computation of predictive means and delta neighborhoods. The predictive means for item respondents and item nonrespondents are calculated using the model coefficients. Then those item respondents whose predictive means are determined to be "close" (based on a distance function taking values within delta) to the item nonrespondents are considered part of the "delta" neighborhood.

Step 5: Assignment of imputed values using a univariate predictive mean. Using a simple random draw from the neighborhood developed in Step 4, a donor is chosen for each item nonrespondent.

If the variables for which Steps 2 through 5 have been completed are part of a complete multivariate set for which multivariate imputation is to be applied, Step 6 is the next step in the process. If the variables for which Steps 2 through 5 are completed are not part of a complete multivariate set, and other variables are still to be imputed, Step 2 is the next step. Otherwise, the process is finished.

Step 6: Determination of multivariate predictive mean neighborhood and assignment of imputed values. With multivariate imputation, the neighborhood is defined based on a vector of predictive means, rather than from a single predictive mean as in the univariate case.

The PMN methodology addresses all of the shortcomings of the unweighted sequential hot-deck method and was widely used for the imputation of a variety of variables in NSDUH, including both continuous and categorical variables with one or more levels. The models were fit using standard modeling procedures in SAS and SUDAAN®, while SAS macros were used to implement the hot-deck step, including the restrictions on the neighborhoods. Although creating a different neighborhood for each item nonrespondent was computationally intensive, the method was implemented successfully. For more details on PMN, see Frechtel et al. (2014).

Appendix D: Generalized Exponential Model Summary

Appendix D: Generalized Exponential Model Summary

This appendix summarizes each model group throughout all stages of modeling the weight calibrations. Unlike much of the other information presented in this report, this appendix provides a model-specific overview of weight calibration, as opposed to a State- or domain-specific one.

The modeling for the 2012 National Survey on Drug Use and Health (NSDUH) involved taking nine generalized exponential model (GEM) groups through five adjustment steps: (1) dwelling unit (DU)–level nonresponse adjustment, (2) DU-level poststratification, (3) selected person-level poststratification, (4) person-level nonresponse adjustment, and (5) respondent person-level poststratification. The sampling weights after DU-level poststratification for this year were reasonably distributed and did not require the additional treatment of the extreme weight adjustment step at the DU level. Because the adaptive fitting strategy for choosing bounds introduced this year does not require the bounds to be as tight as possible (see Section 4.5), an extreme weight adjustment step was performed after respondent person-level poststratification to further control the extreme weight. See Table D for a summary of the distributions of each of the weight components at the national level.

Model-specific summary statistics are shown in Tables D.1a and D.1b to D.9a and D.9b. Included in these tables, for each stage of modeling, are the following: the number of effects that were controlled directly; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for GEM; weighted, unweighted, and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate measure of variance and establishes how much impact a particular stage of modeling has on the distribution of the new product of weights. For more details on bounds, see Section 4.2. At each stage in the modeling, these summary statistics were calculated and used to evaluate the model that was constructed and its corresponding product of weights.

Such circumstances as small sample sizes and exact linear combinations (i.e., singularities) in the realized data led to situations where finalizing models with the originally proposed set of covariates was not possible. The text and exhibits in Sections D.1 to D.9 summarize the decisions made regarding final covariates that were included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit D.1, and for the list of realized final model covariates, see Exhibits D1.1 through D9.5. The following sections establish a series of guidelines to assist in the interpretation of the covariates.

Table D Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (United States)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	30	0.41	40	0.08	9	1.01	19	0.07	4	0.22	4	0.06	1
1%	54	1.00	67	0.51	72	1.01	104	0.34	88	1.00	95	0.20	76
5%	89	1.03	98	0.78	111	1.01	209	0.64	194	1.00	225	0.46	194
10%	132	1.05	147	0.89	162	1.01	340	0.74	322	1.03	368	0.80	322
25%	377	1.08	401	0.99	417	1.15	677	0.87	650	1.11	757	0.96	726
Median	520	1.12	602	1.08	644	1.47	1,334	0.99	1,330	1.22	1,556	1.02	1,549
75%	822	1.19	955	1.19	1,043	5.79	3,483	1.12	3,466	1.38	4,107	1.08	4,069
90%	1,225	1.30	1,393	1.34	1,506	10.61	7,678	1.28	7,716	1.59	10,034	1.23	10,009
95%	1,413	1.40	1,637	1.50	1,829	11.84	11,030	1.42	11,353	1.78	15,350	1.37	15,405
99%	1,694	1.87	1,981	2.16	2,510	13.38	19,642	1.93	19,893	2.48	28,234	1.91	29,312
Maximum	5,182	17.03	6,317	6.33	9,546	35.28	75,644	10.93	83,336	15.95	112,786	5.11	133,926
n	178,586	153,873	153,873	153,858	153,858	87,656	87,656	87,656	87,656	68,309	68,309	68,309	68,309
Max/Mean	8.53	-	8.96	-	12.33	-	25.48	-	28.09	-	29.63	-	35.18

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.
Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

D.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for variable levels are established in Exhibit 3.1 in Chapter 3 (included here as Exhibit D.1 for easy reference). There, a complete list is provided of all variables and associated levels used at any stage of modeling. In this report, each level of a variable is referred to as a covariate. Note that (1) not all variables or levels are present in all stages of modeling; (2) the initial set of covariates, allowing for differences in States across model groups, is the same for all model groups within a stage of modeling; and (3) the initial set of covariates changes across the stages of modeling. Exhibits D.2 through D.5 provide the initial covariates for the stages of modeling, and Exhibits D1.1 through D9.5 provide lists of both the proposed and the final covariates for the nine model groups. This last group of exhibits is grouped by model groups and contains one exhibit for each stage of weight adjustment. The initial variables are found in the "Proposed" column, and the realized covariates are found in the "Final" column.

Section D.3 explains how to create cross-classification tables, which help to illustrate what covariates are controlled for at each stage of the modeling. The general pattern is as follows: directions to follow, semicolon, reason for the change. Sections D.2 and D.3 explain how to use various exhibits for selected model variables to construct these tables. For greater detail on why variable levels are collapsed or dropped, see Section 4.7.

Exhibit D.1 Definition of Levels for Variables

Age (years)
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1,4]

Gender
1: Male, 2: Female[1]

Group Quarters Indicator
1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Hispanicity
1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)
1: 50% - 100%,[1] 2: 10% - > 50%, 3: 0 - >10%

Percentage of Segments That Are Black or African American
1: 50% - 100%, 2: 10% - >50%, 3: 0 - >10%[1]

Percentage of Segments That Are Hispanic or Latino
1: 50% - 100%, 2: 10% - >50%, 3: 0 - >10%[1]

Population Density
1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race (3 levels)
1: White,[1] 2: Black or African American, 3: Other

Race (5 levels)
1: White,[1] 2: Black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: Two or More Races

Relation to Householder
1: Householder or Spouse,[1] 2: Child, 3: Other Relative, 4: Nonrelative

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

States[3]

Model Group 1:	1: Connecticut, 2: Maine, 3: New Hampshire, 4: Rhode Island, 5: Vermont, 6: Massachusetts[1]
Model Group 2:	1: New Jersey,[1] 2: New York, 3: Pennsylvania
Model Group 3:	1: Illinois, 2: Indiana,[1] 3: Michigan, 4: Wisconsin, 5: Ohio
Model Group 4:	1: Iowa, 2: Kansas, 3: Minnesota, 4: Missouri,[1] 5: Nebraska, 6: South Dakota, 7: North Dakota
Model Group 5:	1: Delaware, 2: District of Columbia, 3: Georgia,[1] 4: Maryland, 5: North Carolina, 6: South Carolina, 7: Virginia, 8: West Virginia, 9: Florida
Model Group 6:	1: Alabama, 2: Kentucky, 3: Mississippi, 4: Tennessee[1]
Model Group 7:	1: Arkansas,[1] 2: Louisiana, 3: Oklahoma, 4: Texas
Model Group 8:	1: Colorado, 2: Idaho, 3: Montana, 4: Nevada, 5: New Mexico, 6: Utah, 7: Wyoming, 8: Arizona[1]
Model Group 9:	1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington,[1] 5: California

MSA = metropolitan statistical area.

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-Combined Median Rent and Housing Value (also known as the Socioeconomic Status [SES] indicator) is a composite measure based on rent, housing value, and percent owner occupied.

[3] The States or district assigned to a particular model are based on census divisions.

[4] The age group 50+ was further broken down into 50-64 and 65+ for Person-Level Poststratification Adjustment and Person-Level Extreme Weight Adjustment.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

D.2 Glossary of Terms Used in the Exhibits and Descriptions of the Variables in the Final Model

This glossary provides a list of general terms. Certain other specific terms are sometimes used within a particular section.

All levels present. All levels of the variable under consideration were included in the final model.

Coll. Collapse (levels). These levels of the factor effect were collapsed together. Levels that have been collapsed together no longer appear in the model as separate variables, but rather manifest themselves jointly in the model.

Conv. If model is not convergent, dropping or collapsing of variables is performed.

Drop all levels. All levels of a factor effect were completely removed from the model, as well as any combinations involving this factor.

Drop *level(s)*. These levels of a factor effect were collapsed into the reference set. The dropped levels manifest themselves jointly with the appropriate reference levels.

Drop *level(s)*; singularity/zero sample. During the modeling process, the levels of factor effect(s) listed were removed from the model because of either singularities or sample sizes of zero.

Drop or collapse using *. The asterisk is used as a wildcard character to indicate all levels of that factor effect.

Factor effects. Another name for covariates, or variables, such as "Age." In addition to one-factor effects, two-, and three-factor effects also are referenced, such as "Age × Race" and "Age × Race × Gender."

Hier. Factor effects collapsed/dropped at lower order and the hierarchical effect carries up. This indicates that one or more levels of factor effects were collapsed/dropped in an earlier stage, and that the same action (collapse/drop) was performed on the corresponding levels in all higher-order factor effects containing the dropped/collapsed levels.

Keep *level(s)*. These levels of the factor effect were kept in the model and the remainder into the reference set.

Reference/reference set. The reference levels of factor effects (see Exhibit D.1) are not explicitly listed in the set of model variables, but are represented implicitly in the model in the intercept term. These include one-, two-, and three-factor effects.

Repeat or **Do the same for *(effects)*.** The previous action was repeated for all effect levels listed.

Sing. Singularity is the linear dependence of columns of realized values of the predictors in the model. Any variable that is a linear combination of other variables is either dropped from the model or collapsed with other variables.

D.3 How to Interpret Collapsing and Dropping of Factor Effects

To help visualize what effects were directly controlled for in the model, a table that reflects the collapsing scheme employed can be constructed. The following is a complex example from the 2004 modeling, which demonstrates how to use the information found in Exhibits D1.1 through D9.5.

1. Consider the following entry for the factor effect of State × Age × Race (3 levels), for Model Group 9, for the Person-Level Nonresponse Adjustment.

Three-Factor Effects	Comments
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all age levels in State (2); hier. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all age levels in State (5).

2. Determine the initial range of possible levels for the variables by referring to the variable definitions shown in Exhibit D.1:

State (for the model group in question, in this case, Model Group 9)

Model Group 9: 1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington,[1] 5: California

Age (years)

1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1]

Race (3 levels)

1: White,[1] 2: Black or African American, 3: Other

3. Construct the cross-classification table.

For example, Race (5 levels) is defined this way:

Race (5 Levels)	White	Black or African American	Asian	American Indian or Alaska Native	Two or More Races

Shading indicates the reference-level set.

[1] This is the reference level for this variable. This is the level against which effects of other factor levels are measured.

This is the cross-classification table for State × Race (5 levels):

State × Race (5 levels)	White	Black or African American	Asian	American Indian or Alaska Native	Two or More Races
AK					
HI					
OR					
WA					
CA					

Shading indicates the reference-level set.

The cross-classification table of interest [State × Age × Race (3 levels)] is as follows:

State × Age × Race (3 Levels)	White	Black or African American	Other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Shading indicates the reference-level set.

The number of respondents in that class at this stage of modeling would appear within each cell of the table. Construction of the other cross-classification tables follows the same logic and is only necessary to the point of providing an understanding of the final table.

4. Use the information under the "Final" column definition to determine the combination of factors controlled.

Hier. This means the factor effect was collapsed at a lower order. Because this note is present, examine the information on lower-order factor effects that are the components of the interaction term, State × Race (3 levels) × Age; that is, look at the one-factor and two-factor effects for State, Race (5 levels), and Age, and their accompanying information:

One-Factor Effects	Comments
State	All levels present.
Race (5 Levels)	All levels present.
Age	All levels present.

Two-Factor Effects	Comments
State × Age	All levels present.
State × Race (5 Levels)	Coll. (1,3) & (1,4). Do the same for all other States except (2). Coll. (2,2), (2,3), & (2,4).
Age × Race (3 Levels)	All levels present.

Following these directions, the resulting two-factor table is:

State × Race (5 Levels)	**White**	**Black or African American**	**Asian**	**American Indian or Alaska Native**	**Two or More Races**
AK					
HI					
OR					
WA					
CA					

Shading indicates the reference-level set.

Continuing on to the three-factor level for the same example:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all age levels in State (2); hier. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all age levels in State (5).

The reason for the note "Hier." in the three-factor effects is that collapsing was done on the two-factor interaction term State × Race (5 levels). Because collapsing was done on this term, all three-factor crosses involving State × Race must maintain this same collapsing scheme.

After following the directions, the cross-classification table should appear as follows:

State × Age × Race (3 Levels)	**White**	**Black or African American**	**Other**
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Shading indicates the reference-level set.

The unshaded cells represent the factors directly controlled for by the model (i.e., those factors that were not collapsed or dropped). The shaded cells represent the composite reference set, whose values may be obtained by utilizing the marginal sums, although when changes to the initially proposed set occur, it can make certain reference cell counts indistinguishable.

Exhibit D.2 Covariates for 2012 NSDUH Person Weights (res.sdu.nr)

Variables	Levels	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Population Density	4	3
Group Quarter	3	2
% Black or African American	3	2
% Hispanic or Latino	3	2
% Owner-Occupied	3	2
Rent/Housing	5	4
Two-Factor Effects		
% Owner-Occupied × % Black or African American	3 × 3	4
% Owner-Occupied × % Hispanic or Latino	3 × 3	4
% Owner-Occupied × Rent/Housing	3 × 5	8
Rent/Housing × % Black of African American	3 × 5	8
Rent/Housing × % Hispanic or Latino	3 × 5	8
State × Quarter	Model Specific	
State × Population Density	Model Specific	
State × Group Quarter	Model Specific	
State × % Black or African American	Model Specific	
State × % Hispanic or Latino	Model Specific	
State × % Owner-Occupied	Model Specific	
State × Rent/Housing	Model Specific	
Three-Factor Effects		
State × % Owner-Occupied × % Black or African American	Model Specific	
State × % Owner-Occupied × % Hispanic or Latino	Model Specific	
State × % Owner-Occupied × Rent/Housing	Model Specific	
State × Rent/Housing × % Black or African American	Model Specific	
State × Rent/Housing × % Hispanic or Latino	Model Specific	

Exhibit D.3 Covariates for 2012 NSDUH Person Weights (res.sdu.ps)

Variables	Levels	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Age	5	4
Race (5 levels)	5	4
Gender	2	1
Hispanicity	2	1
Two-Factor Effects		
Age × Race (3 levels)	5 × 3	8
Age × Hispanicity	5 × 2	4
Age × Gender	5 × 2	4
Race (3 levels) × Hispanicity	3 × 2	2
Race (3 levels) × Gender	3 × 2	2
Hispanicity × Gender	3 × 2	1
State × Quarter	Model Specific	
State × Age	Model Specific	
State × Race (5 levels)	Model Specific	
State × Hispanicity	Model Specific	
State × Gender	Model Specific	
Three-Factor Effects		
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8
Age × Race (3 levels) × Gender	5 × 3 × 2	8
Age × Hispanicity × Gender	5 × 2 × 2	4
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2
State × Age × Race (3 levels)	Model Specific	
State × Age × Hispanicity	Model Specific	
State × Age × Gender	Model Specific	
State × Race (3 levels) × Hispanicity	Model Specific	
State × Race (3 levels) × Gender	Model Specific	
State × Hispanicity × Gender	Model Specific	

Exhibit D.4 Covariates for 2012 NSDUH Person Weights (sel.per.ps and res.per.nr)

Variables	Levels	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Age	5	4
Race (5 levels)	5	4
Gender	2	1
Hispanicity	2	1
Relation to Householder	4	3
Population Density	4	3
Group Quarter	3	2
% Black or African American	3	2
% Hispanic or Latino	3	2
% Owner-Occupied	2	2
Rent/Housing	5	4
Two-Factor Effects		
Age × Race (3 levels)	5 × 3	8
Age × Hispanicity	5 × 2	4
Age × Gender	5 × 2	4
Race (3 levels) × Hispanicity	3 × 2	2
Race (3 levels) × Gender	3 × 2	2
Hispanicity × Gender	2 × 2	1
% Owner-Occupied × % Black or African American	3 × 3	4
% Owner-Occupied × % Hispanicity	3 × 3	4
% Owner-Occupied × Rent/Housing	3 × 5	8
Rent/Housing × % Black or African American	3 × 5	8
Rent/Housing × % Hispanic or Latino	3 × 5	8
State × Quarter	Model Specific	
State × Age	Model Specific	
State × Race (5 levels)	Model Specific	
State × Hispanicity	Model Specific	
State × Gender	Model Specific	
State × % Black or African American	Model Specific	
State × % Hispanic or Latino	Model Specific	
State × % Owner-Occupied	Model Specific	
State × Rent/Housing	Model Specific	
Three-Factor Effects		
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8
Age × Race (3 levels) × Gender	5 × 3 × 2	8
Age × Hispanicity × Gender	5 × 2 × 2	4
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2
State × Age × Race (3 levels)	Model Specific	
State × Age × Hispanicity	Model Specific	
State × Age × Gender	Model Specific	
State × Race (3 levels) × Hispanicity	Model Specific	
State × Race (3 levels) × Gender	Model Specific	
State × Hispanicity × Gender	Model Specific	

Exhibit D.5 Covariates for 2012 NSDUH Person Weights (res.per.ps and res.per.ev)

Variables	Levels	Proposed
One-Factor Effects		
Intercept	1	1
State	Model Specific	
Quarter	4	3
Age	6	5
Race (5 levels)	5	4
Gender	2	1
Hispanicity	2	1
Two-Factor Effects		
Age × Race (3 levels)	6 × 3	10
Age × Hispanicity	6 × 2	5
Age × Gender	6 × 2	5
Race (3 levels) × Hispanicity	3 × 2	2
Race (3 levels) × Gender	3 × 2	2
Hispanicity × Gender	2 × 2	1
State × Quarter	Model Specific	
State × Age	Model Specific	
State × Race (5 levels)	Model Specific	
State × Hispanicity	Model Specific	
State × Gender	Model Specific	
Three-Factor Effects		
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10
Age × Race (3 levels) × Gender	6 × 3 × 2	10
Age × Hispanicity × Gender	6 × 2 × 2	5
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2
State × Age × Race (3 levels)	Model Specific	
State × Age × Hispanicity	Model Specific	
State × Age × Gender	Model Specific	
State × Race (3 levels) × Hispanicity	Model Specific	
State × Race (3 levels) × Gender	Model Specific	
State × Hispanicity × Gender	Model Specific	

Appendix D1: Model Group 1: New England

(Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, and Vermont)

Table D.1a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 1: New England)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	5.67	4.48	0.33	1.68366	306	(1.07, 1.50)	(1.07, 1.50)
	5.29	6.08	0.81	1.77527	141	(1.00, 3.57)	(1.00, 3.56)
						(1.10, 1.95)	(1.11, 1.95)
res.sdu.ps	5.29	6.08	0.81	1.77519	232	(0.20, 1.10)	(0.20, 1.10)
	3.23	5.54	1.33	1.95096	222	(0.20, 5.00)	(0.20, 5.00)
						(0.90, 3.15)	(0.90, 3.15)
sel.per.ps	4.15	7.28	2.04	3.72981	332	(0.20, 2.80)	(0.20, 2.80)
	2.24	6.60	1.75	3.86735	277	(0.20, 5.00)	(0.20, 5.00)
						(0.70, 3.58)	(0.70, 3.58)
res.per.nr	2.62	8.60	1.96	3.92329	332	(1.00, 2.90)	(1.00, 2.90)
	2.08	6.73	1.72	4.43213	220	(1.00, 5.00)	(1.00, 5.00)
						(1.10, 2.37)	(1.10, 2.37)
res.per.ps	2.08	6.41	1.73	4.43213	267	(0.20, 1.20)	(0.20, 1.20)
	1.26	2.83	0.51	4.34446	185	(0.20, 5.00)	(0.20, 5.00)
						(0.90, 1.15)	(0.90, 1.15)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.1b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 1: New England)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	66	0.71	70	0.08	18	1.01	19	0.08	8	0.30	8	0.10	3
1%	78	1.00	84	0.30	58	1.01	77	0.28	46	0.84	44	0.20	30
5%	80	1.05	91	0.67	89	1.01	126	0.56	111	1.00	126	0.43	117
10%	88	1.07	100	0.88	103	1.01	167	0.71	156	1.00	179	0.74	164
25%	167	1.09	179	1.00	186	1.14	270	0.84	266	1.06	310	0.95	301
Median	192	1.12	213	1.08	233	1.39	720	0.97	694	1.19	808	1.02	756
75%	453	1.17	533	1.19	577	6.09	1,673	1.14	1,741	1.38	2,168	1.09	2,225
90%	878	1.23	991	1.30	1,121	10.90	4,373	1.34	4,185	1.68	5,301	1.19	5,325
95%	890	1.26	1,090	1.46	1,231	13.10	7,381	1.56	7,332	1.95	9,321	1.44	9,421
99%	907	1.82	1,148	2.20	1,587	15.60	13,779	2.61	13,608	3.00	21,259	2.95	21,451
Maximum	1,456	3.56	2,213	5.00	5,531	27.54	38,702	10.93	47,604	15.95	65,152	5.00	46,935
n	15,380	13,365	13,365	13,363	13,363	7,174	7,174	7,174	7,174	5,615	5,615	5,615	5,615
Max/Mean	4.30	-	5.69	-	12.82	-	22.06	-	27.51	-	29.47	-	21.23

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.
Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 1 Overview

Dwelling Unit Nonresponse

Out of 24 proposed one-factor effects, 23 were included in the model. Variable collapsing was present in the Group Quarter main effect.

For the two-factor effects, variable collapsing or dropping was present in all factors except the percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Hispanic or Latino, State × Group Quarter, State × percent Owner-Occupied, and State × Rent/Housing interactions. Out of 122 proposed variables, 85 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 160 proposed variables, 33 were included in the model.

In the final model, a total of 141 variables were included; see Exhibit D1.1.

Dwelling Unit Poststratification

All 19 proposed one-factor effects were included in the model.

All 86 proposed two-factor effects were included in the model.

For the three-factor effects, variable collapsing or dropping was present in the State × Age × Race and State × Race × Hispanicity interactions. Out of 127 proposed variables, 117 were included in the model.

In the final model, a total of 222 variables were included; see Exhibit D1.2.

Selected Person-Level Poststratification

All 37 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Black or African American, percent Owner-Occupied × percent Hispanic or Latino, Rent/Housing × percent Black or African American, State × Race, State × percent Black or African American, and State × percent Hispanic or Latino interactions. Out of 168 proposed variables, 150 were included in the model.

For the three-factor effects, variable collapsing or dropping was present in all factors except the Age × Race × Gender, Age × Hispanicity × Gender, Race × Hispanicity × Gender, and State × Age × Gender interactions. Out of 127 proposed variables, 90 were included in the model.

In the final model, a total of 277 variables were included; see Exhibit D1.3.

Respondent Person-Level Nonresponse

All 37 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Black or African American, percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Black or African American, State × Race, State × percent Black or African American, State × percent Hispanic or Latino, and State × percent Owner-Occupied interactions. Out of 168 proposed variables, 143 were included in the model.

For the three-factor effects, variable dropping was present in all three-way interactions except State × Age × Gender and State × Hispanicity × Gender. Out of 127 proposed variables, 40 were included in the model.

In the final model, a total of 220 variables were included; see Exhibit D1.4.

Respondent Person-Level Poststratification

All 20 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing was present in the State × Race interactions. Out of 95 proposed two-factor effects, 88 were included in the model.

For the three-factor effects, variable collapsing or dropping was present in all interactions except State × Age × Gender and State × Hispanicity × Gender. Out of 152 proposed variables, 77 were included in the model.

In the final model, a total of 185 variables were included; see Exhibit D1.5.

Exhibit D1.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 1: New England

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**24**	**23**	
Intercept	1	1	1	All levels present.
State	6	5	5	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	1	Coll. (1,2); conv.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**122**	**85**	
% Owner-Occupied × % Black or African American	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (2,1), (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	6 × 4	15	15	All levels present.
State × Population Density	6 × 4	15	5	Keep (1,1), (2,2/3), 3(2,2/3). Drop all others; zero/sing.
State × Group Quarter	6 × 3	10	0	Drop all; conv.
State × % Black or African American	6 × 3	10	2	Keep (1/4,2). Drop all others; zero/sing.
State × % Hispanic or Latino	6 × 3	10	5	Keep (1,1/2), (3,2), (4,1/2). Drop all others; zero/sing.
State × % Owner-Occupied	6 × 3	10	10	All levels present.
State × Rent/Housing	6 × 5	20	20	All levels present.
Three-Factor Effects		**160**	**33**	
State × % Owner-Occupied × % Black or African American	6 × 3 × 3	20	3	Keep (1,2/3,2), (4,2,2). Drop all others; hier./zero/sing.
State × % Owner-Occupied × % Hispanic or Latino	6 × 3 × 3	20	3	Keep (1,2,2), (3,3,2),(4,2,2). Drop all others; hier./zero/sing.
State × % Owner-Occupied × Rent/Housing	6 × 3 × 5	40	18	Keep (1,3,1), (1,2,2/4), Coll. (1,2,3) & (1,3,3). Keep (2,3,2), (2,2,1/2), (3,2,2/3), (4,3,2), (4,2,1/2/3/4), (5,2,1/2/3/4). Drop all others; zero/sing.
State × Rent/Housing × % Black or African American	6 × 3 × 5	40	3	Keep (1,2/3/4,2). Drop all others; hier./zero/sing.
State × Rent/Housing × % Hispanic or Latino	6 × 3 × 5	40	6	Keep (1,2,2), (1,3,1/2), (1,4,2), (3,3,2), (4,2,2). Drop all others; hier./zero/sing.
Total		**306**	**141**	

Exhibit D1.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 1: New England

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**19**	**19**	All levels present.
Intercept	1	1	1	All levels present.
State	6	5	5	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**86**	**86**	All levels present.
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	6 × 4	15	15	All levels present.
State × Age	6 × 5	20	20	All levels present.
State × Race (5 levels)	6 × 5	20	20	All levels present.
State × Hispanicity	6 × 2	5	5	All levels present.
State × Gender	6 × 2	5	5	All levels present.
Three-Factor Effects		**127**	**117**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	6 × 5 × 3	40	36	Coll. (5,1,2) & (5,1,3), repeat for all age levels; conv.
State × Age × Hispanicity	6 × 5 × 2	20	20	All levels present.
State × Age × Gender	6 × 5 × 2	20	20	All levels present.
State × Race (3 levels) × Hispanicity	6 × 3 × 2	10	4	Coll. (1,2,1) & (1,3,1); repeat for all States; drop State VT; conv.
State × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
State × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Total		**232**	**222**	

Exhibit D1.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 1: New England

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**37**	**37**	
Intercept	1	1	1	All levels present.
State	6	5	5	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**168**	**150**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (2/4,1);sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	6 × 4	15	15	All levels present.
State × Age	6 × 5	20	20	All levels present.
State × Race (5 levels)	6 × 5	20	19	Coll. (4,3) & (4,4); conv.
State × Hispanicity	6 × 2	5	5	All levels present.
State × Gender	6 × 2	5	5	All levels present.
State × % Black or African American	6 × 3	10	2	Keep (1/4, 2), drop others; zero.
State × % Hispanic or Latino	6 × 3	10	5	Drop (2/5,1/2), (3,1); zero.
State × % Owner-Occupied	6 × 3	10	10	All levels present.
State × Rent/Housing	6 × 5	20	20	All levels present.
Three-Factor Effects		**127**	**90**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	3	Coll. (1,2,1) & (1,3,1), repeat for all ages and drop age=4; conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	6 × 5 × 3	40	26	Coll. (2,1,2) & (2,1,3), repeat for all age levels and States NH and VT. Drop (2,3/4,2/3); zero, sing, conv.
State × Age × Hispanicity	6 × 5 × 2	20	14	Drop (2,*,1), (5,3/4,1); sing. Conv.
State × Age × Gender	6 × 5 × 2	20	20	All levels present.
State × Race (3 levels) × Hispanicity	6 × 3 × 2	10	0	Coll. (1,2,1) & (1,3,1), repeat for all States, drop all; zero, conv.
State × Race (3 levels) × Gender	6 × 3 × 2	10	9	Coll. (2,2,1) & (2,3,1); conv.
State × Hispanicity × Gender	6 × 2 × 2	5	4	Drop (5,1,1); conv.
Total		**332**	**277**	

Exhibit D1.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 1: New England

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**37**	**37**	
Intercept	1	1	1	All levels present.
State	6	5	5	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**168**	**143**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	3	Drop (2,1); sing.
% Owner-Occupied × Rent/Housing	3 × 5	8	7	Coll. (3,1) & (3,2); conv.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (2/4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	6 × 4	15	15	All levels present.
State × Age	6 × 5	20	20	All levels present.
State × Race (5 levels)	6 × 5	20	18	Coll. (3,3) & (3,4), repeat for State VT; conv.
State × Hispanicity	6 × 2	5	5	All levels present.
State × Gender	6 × 2	5	5	All levels present.
State × % Black or African American	6 × 3	10	2	Keep (1/4,2), drop all others; zero, sing.
State × % Hispanic or Latino	6 × 3	10	5	Drop (2/3/5,1), (2/5,2); zero.
State × % Owner-Occupied	6 × 3	10	5	Coll. (1,2) & (1,3), repeat for all States; conv.
State × Rent/Housing	6 × 5	20	20	All levels present.
Three-Factor-Effects		**127**	**40**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	2	Coll. (1,2,1) & (1,3,1), repeat for age level 2, drop all others; sing./conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	2	Coll. (1,2,1) & (1,3,1), repeat for age level 2, drop all others; conv.
Age × Hispanicity × Gender	5 × 2 × 2	4	3	Drop (4,1,1); conv.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	0	Drop all; conv.
State × Age × Race (3 levels)	6 × 5 × 3	40	0	Drop all; zero/sing./conv.
State × Age × Hispanicity	6 × 5 × 2	20	8	Keep (1/3/4,1/2,1), (2/5,1,1), drop all others; sing./conv.
State × Age × Gender	5 × 5 × 2	20	20	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	10	0	Drop all; zero/conv.
State × Race (3 levels) × Gender	5 × 3 × 2	10	0	Drop all; conv.
State × Hispanicity × Gender	5 × 2 × 2	5	5	All levels present.
Total		**332**	**220**	

Exhibit D1.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 1: New England

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**20**	**20**	
Intercept	1	1	1	All levels present.
State	6	5	5	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**95**	**88**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	6 × 4	15	15	All levels present.
State × Age	6 × 6	25	25	All levels present.
State × Race (5 levels)	6 × 5	20	13	Coll. (1,3) & (1,4), repeat for States NH, RI, and VT, Coll. (2,2) & (2,3) & (2,4) & (2,5); conv.
State × Hispanicity	6 × 2	5	5	All levels present.
State × Gender	6 × 2	5	5	All levels present.
Three-Factor Effects		**152**	**77**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	3	Coll. (1,2,1) & (1,3,1), repeat for ages levels 2 and 3, drop all others; zero/sing./conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	4	Coll. (1,2,1) & (1,3,1), repeat for ages levels 2, 3 and 4, drop all others; conv.
Age × Hispanicity × Gender	6 × 2 × 2	5	4	Drop (5,1,1); conv.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	6 × 5 × 3	50	10	Coll. (1,1,2) & (1,1,3), repeat for age level 2 and all States; heir./conv. Drop all others; zero/sing./conv.
State × Age × Hispanicity	6 × 6 × 2	25	12	Drop (2,3/4/5,1), repeat for States NH, RI, and VT; drop (1,5,1); zero/sing./conv.
State × Age × Gender	6 × 6 × 2	25	25	All levels present.
State × Race (3 levels) × Hispanicity	6 × 3 × 2	10	5	Coll. (1,2,1) & (1,3,1), repeat for all States; heir./zero/conv.
State × Race (3 levels) × Gender	6 × 3 × 2	10	8	Coll. (2,2,1) & (2,3,1) and then drop; hier./conv. Keep others.
State × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Total		**267**	**185**	

Appendix D2: Model Group 2: Middle Atlantic

(New Jersey, New York, and Pennsylvania)

Table D.2a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 2: Middle Atlantic)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	2.91	2.95	0.17	1.19798	153	(1.13, 2.10)	(1.13, 2.10)
	1.57	2.95	0.70	1.20322	119	(1.00, 3.90)	(1.00, 3.90)
						(1.20, 4.10)	(1.20, 4.10)
res.sdu.ps	1.57	2.95	0.70	1.20325	127	(0.61, 1.10)	(0.61, 1.10)
	1.31	4.01	1.44	1.32677	127	(0.37, 5.00)	(0.39, 5.00)
						(0.93, 2.69)	(0.94, 2.69)
sel.per.ps	3.17	7.55	2.30	2.52375	197	(0.59, 2.50)	(0.59, 2.50)
	2.10	5.76	1.37	2.64893	196	(0.38, 2.75)	(0.38, 2.69)
						(0.90, 5.00)	(0.90, 0.90)
res.per.nr	2.13	5.60	1.43	2.76300	197	(1.00, 2.80)	(1.01, 2.80)
	1.91	6.93	1.34	3.19473	187	(1.00, 4.95)	(1.00, 4.95)
						(1.40, 1.48)	(1.40, 1.48)
res.per.ps	1.95	7.07	1.53	3.19473	147	(0.20, 1.30)	(0.20, 1.30)
	0.94	4.81	0.66	3.31270	139	(0.14, 3.26)	(0.14, 3.21)
						(0.90, 1.05)	(1.05, 1.05)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.2b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 2: Middle Atlantic)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	48	0.65	149	0.24	240	1.01	247	0.23	124	0.41	141	0.11	23
1%	383	1.03	422	0.59	374	1.01	392	0.52	325	1.00	370	0.14	109
5%	388	1.07	438	0.76	424	1.01	498	0.67	469	1.03	544	0.24	448
10%	394	1.09	477	0.87	468	1.01	583	0.77	556	1.07	648	0.82	617
25%	481	1.11	554	0.97	565	1.17	800	0.89	779	1.18	927	0.96	926
Median	531	1.19	643	1.04	678	1.55	1,339	0.99	1,364	1.30	1,627	1.03	1,666
75%	565	1.33	827	1.11	865	6.12	4,221	1.09	4,150	1.46	5,204	1.10	5,048
90%	794	1.59	1,237	1.22	1,314	11.60	7,668	1.24	7,464	1.70	10,556	1.24	10,774
95%	1,142	1.90	1,554	1.37	1,737	12.40	9,667	1.37	10,545	1.88	15,250	1.61	15,957
99%	1,628	2.87	1,948	2.25	2,520	12.51	18,676	1.76	19,085	2.56	31,048	1.93	29,857
Maximum	5,182	17.03	6,317	5.00	7,959	26.55	47,814	2.69	56,420	4.95	81,887	3.21	111,268
n	25,030	19,503	19,503	19,500	19,500	11,127	11,127	11,127	11,127	8,158	8,158	8,158	8,158
Max/Mean	8.83	-	8.39	-	9.86	-	15.31	-	18.06	-	19.22	-	26.11

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 2 Overview

Dwelling Unit Nonresponse

All 21 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Population Density and State × Group Quarter interactions. Out of 68 proposed variables, 64 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 64 proposed variables, 34 were included in the model.

In the final model, a total of 119 variables were included; see Exhibit D2.1.

Dwelling Unit Poststratification

All 16 proposed one-factor effects were included in the model.

All 47 proposed two-factor effects were included in the model.

All 64 proposed two-factor effects were included in the model.

In the final model, a total of 127 variables were included; see Exhibit D2.2.

Selected Person-Level Poststratification

All 34 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Rent/Housing interaction. Out of 99 proposed variables, 98 were included in the model.

All 64 proposed two-factor effects were included in the model.

In the final model, a total of 196 variables were included; see Exhibit D2.3.

Respondent Person-Level Nonresponse

All 34 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Race and State × Rent/Housing interactions. Out of 99 proposed variables, 97 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity and State × Race × Hispanicity interactions. Out of 64 proposed variables, 56 were included in the model.

In the final model, a total of 187 variables were included; see Exhibit D2.4.

Respondent Person-Level Poststratification

All 17 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing was present in the State × Race interaction. Out of 53 proposed variables, 52 were included in the model.

For three-factor effects, variable collapsing was present in the Age × Race × Hispanicity and State × Age × Hispanicity interactions. Out of 77 proposed variables, 70 were included in the model.

In the final model, a total of 139 variables were included; see Exhibit D2.5.

Exhibit D2.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 2: Middle Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**21**	**21**	
Intercept	1	1	1	All levels present.
State	3	2	2	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**68**	**64**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	3 × 4	6	6	All levels present.
State × Population Density	3 × 4	6	4	Drop (2.2), (2,3); sing.
State × Group Quarter	3 × 3	4	3	Coll. (2,1) & (2,2); sing.
State × % Black or African American	3 × 3	4	4	All levels present.
State × % Hispanic or Latino	3 × 3	4	4	All levels present.
State × % Owner-Occupied	3 × 3	4	4	All levels present.
State × Rent/Housing	3 × 5	8	7	Drop (2,1); sing.
Three-Factor Effects		**64**	**34**	
State × % Owner-Occupied × % Black or African American	3 × 3 × 3	8	3	Coll. (2,2,1) & (2,2,1) & (2,3,1) & (2,3,2), (3,2,1), (3,2,2) & (3,3,2), drop others; zero./sing./conv.
State × % Owner-Occupied × % Hispanic or Latino	3 × 3 × 3	8	5	Drop (3,3,1), (3,3,2); zero. Coll. (2,3,1) & (2,3,2); conv.
State × % Owner-Occupied × Rent/Housing	3 × 3 × 5	16	10	Keep (2,2,3), (2,2,4), (3,2,*), (3,3,4), coll. (2,2,1) & (2,2,2), (2,3,1) & (2,3,2), (3,3,1) & (3,3,3); sing. Drop others; zero/sing./conv.
State × Rent/Housing × % Black or African American	3 × 3 × 5	16	9	Keep (2,2,1/2), (2,3,*), (2,4,2), (3,1,*), (3,2,*), (3,3,1/2), drop others; zero/sing./conv.
State × Rent/Housing × % Hispanic or Latino	3 × 3 × 5	16	7	Keep (2,2,1/2), (2,3,*), (2,4,*), (3,1,2), (3,2,2), drop others; zero/sing./conv.
Total		**153**	**119**	

Exhibit D2.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 2: Middle Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**16**	**16**	
Intercept	1	1	1	All levels present.
State	3	2	2	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**47**	**47**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	3 × 4	6	6	All levels present.
State × Age	3 × 5	8	8	All levels present.
State × Race (5 levels)	3 × 5	8	8	All levels present.
State × Hispanicity	3 × 2	2	2	All levels present.
State × Gender	3 × 2	2	2	All levels present.
Three-Factor Effects		**64**	**64**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	3 × 5 × 3	16	16	All levels present.
State × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State × Age × Gender	3 × 5 × 2	8	8	All levels present.
State × Race (3 levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
State × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
Total		**127**	**127**	

Exhibit D2.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 2: Middle Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**34**	**34**	
Intercept	1	1	1	All levels present.
State	3	2	2	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**99**	**98**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	3 × 4	6	6	All levels present.
State × Age	3 × 5	8	8	All levels present.
State × Race (5 levels)	3 × 5	8	8	All levels present.
State × Hispanicity	3 × 2	2	2	All levels present.
State × Gender	3 × 2	2	2	All levels present.
State × % Black or African American	3 × 3	4	4	All levels present.
State × % Hispanic or Latino	3 × 3	4	4	All levels present.
State × % Owner-Occupied	3 × 3	4	4	All levels present.
State × Rent/Housing	3 × 5	8	7	Coll. (2.1) & (2.2); sing.
Three-Factor Effects		**64**	**64**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	3 × 5 × 3	16	16	All levels present.
State × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State × Age × Gender	3 × 5 × 2	8	8	All levels present.
State × Race (3 levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
State × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
Total		**197**	**196**	

Exhibit D2.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 2: Middle Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**34**	**34**	
Intercept	1	1	1	All levels present.
State	3	2	2	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**99**	**97**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	3 × 4	6	6	All levels present.
State × Age	3 × 5	8	8	All levels present.
State × Race (5 levels)	3 × 5	8	7	Coll. (2,3) & (2,4); sing.
State × Hispanicity	3 × 2	2	2	All levels present.
State × Gender	3 × 2	2	2	All levels present.
State × % Black or African American	3 × 3	4	4	All levels present.
State × % Hispanic or Latino	3 × 3	4	4	All levels present.
State × % Owner-Occupied	3 × 3	4	4	All levels present.
State × Rent/Housing	3 × 5	8	7	Coll. (2,1) & (2,2); sing.
Three-Factor Effects		**64**	**56**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	2	Coll. (4,2,1) & (4,3,1), repeat for all age levels, drop (3,2/3,1) (4,2/3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	3 × 5 × 3	16	16	All levels present.
State × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State × Age × Gender	3 × 5 × 2	8	8	All levels present.
State × Race (3 levels) × Hispanicity	3 × 3 × 2	4	2	Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); conv.
State × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
State × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
Total		**197**	**187**	

Exhibit D2.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 2: Middle Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**17**	**17**	
Intercept	1	1	1	All levels present.
State	3	2	2	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**53**	**52**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	3 × 4	6	6	All levels present.
State × Age	3 × 6	10	10	All levels present.
State × Race (5 levels)	3 × 5	8	7	Coll. (3,3) & (3,4); conv.
State × Hispanicity	3 × 2	2	2	All levels present.
State × Gender	3 × 2	2	2	All levels present.
Three-Factor Effects		**77**	**70**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	9	Coll. (5,2,1) & (5,3,1); conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race(3 levels)	3 × 6 × 3	20	20	All levels present.
State × Age × Hispanicity	3 × 6 × 2	10	4	Drop (3,3,1), (3,4,1), (3,5,1), repeat for State NY; conv.
State × Age × Gender	3 × 6 × 2	10	10	All levels present.
State × Race (3 levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
State × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
Total		**147**	**139**	

Appendix D3: Model Group 3: East North Central

(Illinois, Indiana, Michigan, Ohio, and Wisconsin)

Table D.3a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 3: East North Central)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	4.31	4.38	0.18	1.20223	255	(1.03, 1.20)	(1.04, 1.20)
	4.39	5.06	0.52	1.18185	181	(1.00, 4.08)	(1.00, 4.08)
						(1.10, 3.36)	(1.10, 3.36)
res.sdu.ps	4.39	5.06	0.52	1.18186	197	(0.72, 1.10)	(0.72, 1.10)
	0.98	1.99	0.39	1.21947	197	(0.25, 4.96)	(0.26, 4.95)
						(0.90, 3.27)	(0.90, 3.27)
sel.per.ps	2.97	5.04	1.14	2.46105	287	(0.20, 3.00)	(0.20, 3.00)
	1.47	3.08	0.50	2.41330	281	(0.20, 4.74)	(0.20, 4.68)
						(0.90, 1.76)	(0.90, 1.76)
res.per.nr	1.44	3.01	0.57	2.46151	287	(1.00, 2.90)	(1.00, 2.89)
	1.04	2.95	0.44	2.72169	269	(1.00, 5.00)	(1.00, 5.00)
						(1.10, 4.45)	(1.10, 4.44)
res.per.ps	1.06	2.92	0.47	2.72169	227	(0.20, 1.10)	(0.20, 1.10)
	0.40	1.43	0.35	2.78599	211	(0.20, 4.34)	(0.20, 4.33)
						(0.90, 1.05)	(0.90, 1.05)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.

[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.3b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 3: East North Central)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	34	0.79	40	0.26	42	1.01	55	0.12	47	0.36	49	0.11	37
1%	302	1.00	345	0.68	325	1.01	348	0.57	327	1.00	377	0.20	166
5%	307	1.04	366	0.89	395	1.01	459	0.75	433	1.03	512	0.72	456
10%	376	1.07	423	0.95	433	1.01	504	0.82	488	1.08	586	0.90	572
25%	399	1.11	453	1.02	478	1.15	641	0.92	627	1.16	760	0.98	765
Median	416	1.14	496	1.07	539	1.38	970	1.00	1,020	1.27	1,236	1.02	1,248
75%	494	1.22	593	1.14	651	5.65	3,125	1.08	3,149	1.38	3,939	1.05	3,954
90%	914	1.33	1,014	1.23	1,109	10.67	6,052	1.22	5,994	1.53	8,032	1.11	8,045
95%	1,117	1.40	1,221	1.31	1,324	11.51	7,171	1.32	7,332	1.66	10,385	1.23	10,498
99%	1,240	1.80	1,390	1.62	1,658	11.91	14,502	1.60	13,979	2.04	19,750	1.56	19,866
Maximum	1,347	6.71	5,490	4.95	6,654	25.29	68,551	4.68	33,793	7.92	47,292	4.33	71,502
n	33,444	28,338	28,338	28,335	28,335	16,573	16,573	16,573	16,573	12,800	12,800	12,800	12,800
Max/Mean	2.70	-	9.34	-	10.40	-	29.14	-	14.43	-	15.59	-	23.58

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 3 Overview

Dwelling Unit Nonresponse

All 23 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in State × Group Quarter and State × percent Hispanic or Latino interactions. Out of 104 proposed variables, 100 were included in the model.

For three-factor effects, variable collapsing or dropping was present in all interactions. Out of 128 proposed variables, 58 were included in the model.

In the final model, a total of 181 variables were included; see Exhibit D3.1.

Dwelling Unit Poststratification

All 18 proposed one-factor effects were included in the model.

All 73 proposed two-factor effects were included in the model.

All 106 proposed three-factor effects were included in the model.

In the final model, a total of 197 variables were included; see Exhibit D3.2.

Selected Person-Level Poststratification

All 36 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × percent Hispanic or Latino interaction. Out of 145 proposed variables, 144 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, State × Age × Race, and State × Race × Hispanicity interactions. Out of 106 proposed variables, 101 were included in the model.

In the final model, a total of 281 variables were included; see Exhibit D3.3.

Respondent Person-Level Nonresponse

All 36 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing and dropping was present in the State × percent Hispanic or Latino interaction. Out of 145 proposed variables, 144 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, State × Age × Race, State × Age × Hispanicity, State × Race × Hispanicity, and State × Race × Gender interactions. Out of 106 proposed variables, 89 were included in the model.

In the final model, a total of 269 variables were included; see Exhibit D3.4.

Respondent Person-Level Poststratification

All 19 proposed one-factor effects were included in the model.

All 81 proposed two-factor effects were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, State × Age × Race, State × Age × Hispanicity, and State × Race × Hispanicity interactions. Out of 127 proposed variables, 111 were included in the model.

In the final model, a total of 211 variables were included; see Exhibit D3.5.

Exhibit D3.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 3: East North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**23**	**23**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**104**	**100**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Population Density	5 × 4	12	12	All levels present.
State × Group Quarter	5 × 3	8	5	Drop (4,2); sing. Coll. (1,1) & (1,2); conv. Drop (1,1/2); conv.
State × % Black or African American	5 × 3	8	8	All levels present.
State × % Hispanic or Latino	5 × 3	8	7	Drop (4,1); sing.
State × % Owner-Occupied	5 × 3	8	8	All levels present.
State × Rent/Housing	5 × 5	16	16	All levels present.
Three-Factor Effects		**128**	**58**	
State × % Owner-Occupied × % Black or African American	5 × 3 × 3	16	9	Drop (4,3,1), (4,2,2), (3,3,1), (4,3,2); sing. Coll. (1,3,1) & (1,3,2), (1,2,1) & (1,2,2); conv. Drop (4,2,1); conv.
State × % Owner-Occupied × % Hispanic or Latino	5 × 3 × 3	16	4	Drop (4,3,1), (4,2,1); hier. Drop (1,3,1), (5,2,1), (3,2,1), (1,3,1), (5,3,1), (5,3,2), (3,3,1), (3,3,2), (4,3,2); sing. Drop (4,2,2); conv.
State × % Owner-Occupied × Rent/Housing	5 × 3 × 5	32	16	Drop (5,3,4), (3,3,2), (3,3,4), (4,3,2), (3,3,1), (4,3,1), (4,3,3), (4,3,4), (4,2,1); sing. Coll. (5,3,1) & (5,3,2), (5,3,1/2) & (5,3,3), (1,3,3) & (1,3,4), (1,2,1) & (1,2,2), (1,2,1/2) & (1,2,3), (5,2,3) & (5,2,4); conv. Drop (3,3,3); conv.
State × Rent/Housing × % Black or African American	5 × 3 × 5	32	20	Drop (5,4,1), (3,3,1), (3,4,1), (3,4,2), (4,2,1), (4,3,2), (4,1,1), (4,1,2), (4,2,2), (4,3,1), (4,4,1), (4,4,2); sing.
State × Rent/Housing × % Hispanic or Latino	5 × 3 × 5	32	9	Drop (4,1,1), (4,2,1), (4,3,1), (4,4,1); hier. Drop (1,3,1), (1,4,1), (5,1,1), (5,3,1), (3,1,1), (3,2,1), (3,3,2), (3,4,2), (4,2,2), (4,3,2), (5,2,1), (5,3,2), (5,4,1), (3,3,1), (3,4,1), (4,4,2); sing. Drop (1,3,2), (5,1,2), (5,4,2); conv.
Total		**255**	**181**	

Exhibit D3.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 3: East North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**18**	**18**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**73**	**73**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
Three-Factor Effects		**106**	**106**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	5 × 5 × 3	32	32	All levels present.
State × Age × Hispanicity	5 × 5 × 2	16	16	All levels present.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**197**	**197**	

Exhibit D3.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 3: East North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**36**	**36**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**145**	**144**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
State × % Black or African American	5 × 3	8	8	All levels present.
State × % Hispanic or Latino	5 × 3	8	7	Drop (4,1); sing.
State × % Owner-Occupied	5 × 3	8	8	All levels present.
State × Rent/Housing	5 × 5	16	16	All levels present.
Three-Factor Effects		**106**	**101**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	7	Coll. (3,2,1) & (3,3,1); sing.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	5 × 5 × 3	32	31	Coll. (4,3,2) & (4,3,3); conv.
State × Age × Hispanicity	5 × 5 × 2	16	16	All levels present.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (4,2,1) & (4,3,1), (3,2,1) & (3,3,1); conv. Drop (3,2/3,1); conv.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**287**	**281**	

Exhibit D3.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 3: East North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**36**	**36**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**145**	**144**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
State × % Black or African American	5 × 3	8	8	All levels present.
State × % Hispanic or Latino	5 × 3	8	7	Drop (4,1); sing.
State × % Owner-Occupied	5 × 3	8	8	All levels present.
State × Rent/Housing	5 × 5	16	16	All levels present.
Three-Factor Effects		**106**	**89**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (3,2,1) & (3,3,1); sing. Coll. (2,2,1) & (2,3,1), (4,2,1) & (4,3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	5 × 5 × 3	32	26	Coll. (4,4,2) & (4,4,3); sing. Coll. (4,3,2) & (4,3,3), (5,4,2) & (5,4,3), (5,3,2) & (5,3,3); conv. Drop (4,3,2/3), (4,4,2/3); conv.
State × Age × Hispanicity	5 × 5 × 2	16	15	Drop (4,4,1); conv.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	2	Coll. (4,2,1) & (4,3,1), (3,2,1) & (3,3,1), (5,2,1) & (5,3,1), (1,2,1) & (1,3,1); conv. Drop (3,2/3,1), (4,2/3,1); conv.
State × Race (3 levels) × Gender	5 × 3 × 2	8	7	Coll. (4,2,1) & (4,3,1); conv.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**287**	**269**	

Exhibit D3.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 3: East North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**19**	**19**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**81**	**81**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 6	20	20	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
Three-Factor Effects		**127**	**111**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	7	Coll. (5,2,1) & (5,3,1); sing. Drop (5,2/3,1); conv. Coll. (4,2,1) & (4,3,1) conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	5 × 6 × 3	40	33	Coll. (4,4,2) & (4,4,3); sing. Drop (4,5,2), (4,5,3); sing. Coll. (5,5,2) & (5,5,3), (1,5,2) & (1,5,3), (1,4,2) & (1,4,3); conv. Drop (5,5,2/3); conv.
State × Age × Hispanicity	5 × 6 × 2	20	16	Drop (4,4,1), (4,5,1); sing. Drop (1,5,1), (3,5,1); conv.
State × Age × Gender	5 × 6 × 2	20	20	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	6	Coll. (4,2,1) & (4,3,1), (3,2,1) & (3,3,1); conv.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**227**	**211**	

Appendix D4: Model Group 4: West North Central

(Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, and South Dakota)

Table D.4a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 4: West North Central)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	2.95	2.66	0.22	1.54552	357	(1.05, 1.20)	(1.06, 1.19)
	0.89	1.04	0.10	1.55753	149	(1.00, 2.58)	(1.00, 2.58)
						(1.05, 1.14)	(1.05, 1.14)
res.sdu.ps	0.89	1.04	0.10	1.55756	267	(0.69, 1.05)	(0.69, 1.05)
	2.37	3.90	0.90	1.56469	255	(0.26, 4.98)	(0.27, 4.97)
						(0.95, 1.11)	(0.95, 1.10)
sel.per.ps	3.26	5.76	1.47	3.06000	377	(0.20, 2.75)	(0.20, 2.75)
	2.26	6.01	1.65	3.37592	315	(0.20, 5.00)	(0.20, 5.00)
						(0.35, 1.91)	(0.35, 1.86)
res.per.nr	2.51	6.54	1.87	3.49120	377	(1.00, 2.95)	(1.00, 2.95)
	2.14	7.14	1.70	3.77739	295	(1.00, 5.00)	(1.00, 5.00)
						(1.20, 1.85)	(1.20, 1.85)
res.per.ps	2.19	7.11	1.69	3.77739	307	(0.20, 2.20)	(0.20, 2.20)
	1.69	6.06	1.66	3.96150	243	(0.20, 5.00)	(0.20, 5.00)
						(0.90, 1.77)	(1.33, 1.77)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.4b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 4: West North Central)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	77	0.85	78	0.27	33	1.01	40	0.08	25	0.41	33	0.09	7
1%	78	1.00	82	0.55	87	1.01	98	0.20	72	0.89	85	0.20	61
5%	83	1.01	87	0.74	101	1.01	154	0.47	128	1.00	148	0.49	134
10%	88	1.03	96	0.83	117	1.01	213	0.66	184	1.01	215	0.82	192
25%	153	1.05	158	0.94	172	1.13	455	0.85	426	1.08	475	0.94	453
Median	458	1.08	490	1.05	504	1.43	957	1.00	985	1.17	1,113	1.00	1,111
75%	816	1.10	886	1.17	869	5.82	2,286	1.16	2,174	1.33	2,547	1.08	2,606
90%	1,054	1.14	1,138	1.35	1,137	10.94	5,769	1.36	5,569	1.54	7,177	1.19	6,970
95%	1,091	1.18	1,193	1.51	1,294	12.13	8,332	1.52	9,098	1.78	11,446	1.35	11,201
99%	1,154	1.26	1,377	2.27	1,775	13.08	14,252	2.36	16,619	2.64	22,355	2.83	23,454
Maximum	2,389	2.58	2,673	4.97	5,701	35.28	40,189	5.00	31,554	5.00	63,117	5.00	55,395
n	15,937	14,672	14,672	14,671	14,671	7,926	7,926	7,926	7,926	6,342	6,342	6,342	6,342
Max/Mean	4.70	-	4.84	-	10.05	-	19.00	-	14.62	-	23.40	-	20.54

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 4 Overview

Dwelling Unit Nonresponse

All 25 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in all effects except percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, State × Quarter, State × percent Owner-Occupied, and State × Rent/Housing. Out of 140 proposed variables, 109 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 192 proposed variables, 15 were included in the model.

In the final model, a total of 149 variables were included; see Exhibit D4.1.

Dwelling Unit Poststratification

All 20 proposed one-factor effects were included in the model.

All 99 proposed two-factor effects were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, Race × Hispanicity × Gender, State × Age × Race, and State × Race × Hispanicity interactions. Out of 148 proposed variables, 136 were included in the model.

In the final model, a total of 255 variables were included; see Exhibit D4.2.

Selected Person-Level Poststratification

All 38 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the Race × Hispanicity, percent Owner-Occupied × percent Black or African American, Rent/Housing × percent Black or African American, Rent/Housing × percent Hispanic or Latino, State × Race, State × percent Black or African American, and State × percent Hispanic or Latino interactions. Out of 191 proposed variables, 171 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, Race × Hispanicity × Gender, State × Age × Race, and State × Race × Hispanicity interactions. Out of 148 proposed variables, 106 were included in the model.

In the final model, a total of 315 variables were included; see Exhibit D4.3.

Respondent Person-Level Nonresponse

All 38 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Black or African American, Rent/Housing × percent Black or African American, Rent/Housing × percent Hispanic or Latino, State × percent Black or African American, and State × percent Hispanic or Latino interactions. Out of 191 proposed variables, 172 were included in the model.

Variable collapsing or dropping was present in all three-factor effects except the Age × Race × Gender, Age × Hispanicity × Gender, and State × Age × Gender interactions. Out of 148 proposed variables, 85 were included in the model.

In the final model, a total of 295 variables were included; see Exhibit D4.4.

Respondent Person-Level Poststratification

All 21 proposed one-factor effects were included in the model.

All 109 proposed two-factor effects were included in the model.

For three-factor effects, all levels were present for the Race × Hispanicity × Gender and State × Age × Gender interactions. All the others were affected by variable collapsing or dropping. Out of 177 proposed variables, 113 were included in the model.

In the final model, a total of 243 variables were included; see Exhibit D4.5.

Exhibit D4.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 4: West North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**25**	**25**	
Intercept	1	1	1	All levels present.
State	7	6	6	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**140**	**109**	
% Owner-Occupied × % Black or African American	3 × 3	4	3	Coll. (3,1) & (3,2); zero.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present
Rent/Housing × % Black or African American	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
State × Quarter	7 × 4	18	18	All levels present.
State × Population Density	7 × 4	18	14	Coll. (1,1) & (1,2), repeat for States 5,6, and 7; zero.
State × Group Quarter	7 × 3	12	2	Coll. (7,1) & (7,2), (5,1) & (5,2), drop rest; conv.
State × % Black or African American	7 × 3	12	7	Coll. (1,1) & (1,2), do the same for States 3,5,6, and 7; zero.
State × % Hispanic or Latino	7 × 3	12	5	Coll. (3,1) & (3,2), (5,1) & (5,2); sing. Coll. (1,1) & (1,2), drop States 6 and 7; zero.
State × % Owner-Occupied	7 × 3	12	12	All levels present.
State × Rent/Housing	7 × 5	24	24	All levels present.
Three-Factor Effects		**192**	**15**	
State × % Owner-Occupied × % Black or African American	7 × 3 × 3	24	0	Drop all; zero/sing./conv.
State × % Owner-Occupied × % Hispanic or Latino	7 × 3 × 3	24	1	Coll. (2,2,1) & (2,2,2), drop rest; conv.
State × % Owner-Occupied × Rent/Housing	7 × 3 × 5	48	11	Coll. (1,2,1) & (1,2,2), repeat for States 3 and 5, keep (1,2,3), (3,2,3), (5,2,3), (5,2,4), (7,2,*), drop rest; conv.
State × Rent/Housing × % Black or African American	7 × 3 × 5	48	2	Coll. (1,1,1) & (1,1,2) & (1,2,1) & (1,2,2), (3,4,1) & (3,4,2), drop rest; conv.
State × Rent/Housing × % Hispanic or Latino	7 × 3 × 5	48	1	Coll. (1,1,1) & (1,1,2) & (1,2,1) & (1,2,2), drop rest; conv.
Total		**357**	**149**	

Exhibit D4.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 4: West North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**20**	**20**	
Intercept	1	1	1	All levels present.
State	7	6	6	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**99**	**99**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	7 × 4	18	18	All levels present.
State × Age	7 × 5	24	24	All levels present.
State × Race (5 levels)	7 × 5	24	24	All levels present.
State × Hispanicity	7 × 2	6	6	All levels present.
State × Gender	7 × 2	6	6	All levels present.
Three-Factor Effects		**148**	**136**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (2,2,1) & (2,3,1); sing. Coll. (1,2,1) & (1,3,1), repeat for age levels 3 and 4; conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,2,1) & (3,2,1); conv.
State × Age × Race (3 levels)	7 × 5 × 3	48	47	Coll. (6,1,2) & (6,1,3); conv.
State × Age × Hispanicity	7 × 5 × 2	24	24	All levels present.
State × Age × Gender	7 × 5 × 2	24	24	All levels present.
State × Race (3 levels) × Hispanicity	7 × 3 × 2	12	6	Coll. (1,2,1) & (1,3,1), (6,2,1) & (6,3,1); zero. Do the same for remaining States; conv.
State × Race (3 levels) × Gender	7 × 3 × 2	12	12	All levels present.
State × Hispanicity × Gender	7 × 2 × 2	6	6	All levels present.
Total		**267**	**255**	

Exhibit D4.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 4: West North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**38**	**38**	
Intercept	1	1	1	All levels present.
State	7	6	6	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**191**	**171**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	1	Coll. (2,1) & (3,1); conv.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	3	Coll. (3,1) & (3,2); zero.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
State × Quarter	7 × 4	18	18	All levels present.
State × Age	7 × 5	24	24	All levels present.
State × Race (5 levels)	7 × 5	24	23	Coll. (1,3) & (1,4); conv.
State × Hispanicity	7 × 2	6	6	All levels present.
State × Gender	7 × 2	6	6	All levels present.
State × % Black or African American	7 × 3	12	7	Coll. (1,1) & (1,2), repeat for States 3,5, and 6; zero. Coll. (5,1) & (5,2); sing.
State × % Hispanic or Latino	7 × 3	12	4	Drop (6/7,1/2), coll. (1,1) & (1,2); zero. Coll. (3,1) & (3,2), (5,1) & (5,2); sing. Coll. (2,1) & (2,2); conv.
State × % Owner-Occupied	7 × 3	12	12	All levels present.
State × Rent/Housing	7 × 5	24	24	All levels present.
Three-Factor Effects		**148**	**106**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (3,2,1) & (3,3,1); zero. Coll. (*,2,1) & (*,3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); zero.
State × Age × Race (3 levels)	7 × 5 × 3	48	23	Coll. (7,4,2) & (7,4,3); sing. Repeat for all States and age levels, drop (5,4,2/3); conv.
State × Age × Hispanicity	7 × 5 × 2	24	22	Drop (3,4,1); sing. Drop (7,4,1); conv.
State × Race (3 levels) × Hispanicity	7 × 3 × 2	12	2	Coll. (3,2,1) & (3,3,1), (6,2,1) & (6,3,1); zero. Drop rest; conv.
State × Race (3 levels) × Gender	7 × 3 × 2	12	12	All levels present.
State × Age × Gender	7 × 5 × 2	24	24	All levels present.
State × Hispanicity × Gender	7 × 2 × 2	6	6	All levels present.
Total		**377**	**315**	

Exhibit D4.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 4: West North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**38**	**38**	
Intercept	1	1	1	All levels present.
State	7	6	6	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**191**	**172**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	3	Coll. (3,1) & (3,2); zero
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	6	Coll. (3,1) & (3,2); zero. Coll. (4,1) & (4,2); sing.
State × Quarter	7 × 4	18	18	All levels present.
State × Age	7 × 5	24	24	All levels present.
State × Race (5 levels)	7 × 5	24	24	All levels present.
State × Hispanicity	7 × 2	6	6	All levels present.
State × Gender	7 × 2	6	6	All levels present.
State × % Black or African American	7 × 3	12	7	Coll. (1,1) & (1,2), repeat for States 3,6, and 7; zero. Coll. (5,1) & (5,2); sing.
State × % Hispanic or Latino	7 × 3	12	3	Coll. (1,1) & (1,2); Drop (6/7,1/2); zero. Coll. (3,1) & (3,2); sing. Coll. (2,1) & (2,2), drop (5, 1/2); conv.
State × % Owner-Occupied	7 × 3	12	12	All levels present.
State × Rent/Housing	7 × 5	24	24	All levels present.
Three-Factor Effects		**148**	**85**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	1	Drop (2/3/4, 2/3, 1), coll. (1,2,1) & (1,3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); zero.
State × Age × Race (3 levels)	7 × 5 × 3	48	17	Coll. (1,1,2) & (1,1,3), repeat for all other States and age levels, drop (1/2/5/6/7, 4, 2/3), (5/6,3,2/3); conv.
State × Age × Hispanicity	7 × 5 × 2	24	15	Drop (3,4,1); sing. Drop (3, 1/2/3, 1), (1/5/6/7,4,1), (6,3,1); conv.
State × Age × Gender	7 × 5 × 2	24	24	All levels present.
State × Race (3 levels) × Hispanicity	7 × 3 × 2	12	3	Coll. (5,2,1) & (5,3,1), repeat for States 6 and 7; zero. Drop (1/2/3, 2/3, 1); conv.
State × Race (3 levels) × Gender	7 × 3 × 2	12	8	Coll. (1,2,1) & (1,3,1), repeat for States 3, 5, and 6; conv.
State × Hispanicity × Gender	7 × 2 × 2	6	4	Drop (3,1,1) & (6,1,1); conv.
Total		**377**	**295**	

Exhibit D4.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 4: West North Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**21**	**21**	
Intercept	1	1	1	All levels present.
State	7	6	6	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**109**	**109**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	7 × 4	18	18	All levels present.
State × Age	7 × 6	30	30	All levels present.
State × Race (5 levels)	7 × 5	24	24	All levels present.
State × Hispanicity	7 × 2	6	6	All levels present.
State × Gender	7 × 2	6	6	All levels present.
Three-Factor Effects		**177**	**113**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	2	Coll. (1,2,2) & (1,2,3), (2,2,2) & (2,2,3), drop rest; conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	7	Coll. (1,4,2) & (1,4,3), drop (1,5,2/3); conv.
Age × Hispanicity × Gender	6 × 2 × 2	5	4	Drop (5,1,1); sing.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	7 × 6 × 3	60	31	Coll. (6,3,2) & (6,3,3), (7,5,2) & (7,5,3); zero. Coll. (3,5,2) & (3,5,3), (5,4,2) & (5,4,3), (6,5,2) & (6,5,3), (7,4,2) & (7,4,3); sing. Coll. (1,1,2) & (1,1,3), repeat for States 2, 3, and 5; conv. Coll. (1,2,2) & (1,2,3), repeat for States 2, 3, and 5; conv. Coll. (1,3,2) & (1,3,3), repeat for States 2, 3, 5, and 7; conv. Coll. (2,4,2) & (2,4,3), (3,4,2) & (3,4,3), drop (1,4/5,2/3), (2,5,2/3), (5,5,2/3); conv.
State × Age × Hispanicity	7 × 6 × 2	30	19	Drop (1/3/7, 5, 1); zero. Drop (2/5/6,5,1), (3/7,4,1); sing. Drop (1/2/6,4,1); conv.
State × Age × Gender	7 × 6 × 2	30	30	All levels present.
State × Race (3 levels) × Hispanicity	7 × 3 × 2	12	3	Coll. (2,2,1) & (2,3,1), (5,2,1) & (5,3,1), drop (3/6/7,2/3,1); conv. Coll. (1,2,1) & (1,3,1); conv.
State × Race (3 levels) × Gender	7 × 3 × 2	12	10	Coll. (1,2,1) & (1,3,1), (5,2,1) & (5,3,1); conv.
State × Hispanicity × Gender	7 × 2 × 2	6	5	Drop (6,1,1); conv.
Total		**307**	**243**	

Appendix D5: Model Group 5: South Atlantic

(Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, and West Virginia)

Table D.5a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 5: South Atlantic)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	4.06	3.61	0.24	1.50780	459	(1.07, 1.74)	(1.08, 1.74)
	2.53	3.03	0.21	1.49510	306	(1.00, 2.30)	(1.00, 2.26)
						(1.00, 2.56)	(1.00, 2.55)
res.sdu.ps	2.53	3.03	0.21	1.49514	337	(0.61, 1.10)	(0.61, 1.10)
	1.67	3.62	0.72	1.56442	333	(0.20, 5.00)	(0.20, 5.00)
						(0.90, 3.43)	(0.90, 3.43)
sel.per.ps	2.68	5.68	1.09	2.79932	467	(0.20, 2.30)	(0.20, 2.30)
	1.34	3.05	0.71	2.87926	440	(0.20, 5.00)	(0.20, 5.00)
						(0.30, 4.18)	(0.30, 4.10)
res.per.nr	1.35	3.06	0.75	2.98016	467	(1.00, 3.00)	(1.00, 3.00)
	1.54	4.26	1.02	3.41176	400	(1.00, 5.00)	(1.00, 5.00)
						(1.00, 5.00)	(1.00, 5.00)
res.per.ps	1.64	4.27	1.04	3.41176	387	(0.20, 1.80)	(0.20, 1.80)
	1.07	3.62	0.88	3.49532	328	(0.30, 5.00)	(0.30, 5.00)
						(0.90, 1.16)	(0.90, 1.16)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.5b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 5: South Atlantic)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	37	0.41	43	0.20	9	1.01	19	0.11	4	0.22	4	0.06	1
1%	38	1.00	47	0.47	55	1.01	73	0.27	55	1.00	59	0.30	49
5%	53	1.04	64	0.81	74	1.01	172	0.63	156	1.00	175	0.50	162
10%	57	1.06	72	0.89	93	1.01	301	0.72	288	1.02	332	0.81	295
25%	295	1.10	327	1.00	323	1.11	870	0.86	814	1.08	886	0.95	853
Median	634	1.15	723	1.09	829	1.52	1,539	0.99	1,537	1.19	1,797	1.02	1,799
75%	916	1.22	1,185	1.22	1,258	6.37	4,902	1.13	4,626	1.35	4,912	1.08	4,809
90%	1,458	1.31	1,643	1.39	1,772	11.35	10,167	1.29	10,643	1.57	13,464	1.20	13,228
95%	1,675	1.38	1,837	1.56	2,113	12.24	13,351	1.45	14,683	1.77	20,543	1.29	20,555
99%	1,733	1.52	2,106	2.21	2,926	14.36	23,414	1.93	22,785	2.84	34,252	1.94	35,521
Maximum	3,566	3.53	4,030	5.00	9,546	21.91	59,251	5.87	83,336	10.45	91,132	5.11	97,222
n	30,681	26,239	26,239	26,236	26,236	13,632	13,632	13,632	13,632	10,883	10,883	10,883	10,883
Max/Mean	5.15	-	4.97	-	10.63	-	16.04	-	22.34	-	19.50	-	20.80

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 5 Overview

Dwelling Unit Nonresponse

All 27 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Population Density, State × Group Quarter, and State × percent Hispanic or Latino interactions. Out of 176 proposed variables, 157 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Many factors were excluded due to zero sample sizes or exact linear combinations. Out of 256 proposed variables, 122 were included in the model.

In the final model, a total of 306 variables were included; see Exhibit D5.1.

Dwelling Unit Poststratification

All 22 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Race interaction. Out of 125 proposed variables, 124 were included in the model.

For three-factor effects, variable collapsing was present in the State × Age × Hispanicity and State × Race × Hispanicity interactions. Out of 190 proposed variables, 187 were included in the model.

In the final model, a total of 333 variables were included; see Exhibit D5.2.

Selected Person-Level Poststratification

All 40 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing and dropping was present in the State × Race and State × percent Hispanic or Latino interactions. Out of 237 proposed variables, 232 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, State × Age × Race, State × Age × Hispanicity, and State × Race × Hispanicity interactions. Out of 190 proposed variables, 168 were included in the model.

In the final model, a total of 440 variables were included; see Exhibit D5.3.

Respondent Person-Level Nonresponse

All 40 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the State × Race, State × percent Hispanic or Latino, and State × Rent/Housing interactions. Out of 237 proposed variables, 226 were included in the model.

For three-factor effects, all levels were present for the Age × Race ×Gender, Age × Hispanicity × Gender, Race × Hispanicity × Gender, State × Age × Gender, and State × Hispanicity × Gender interactions. Out of 190 proposed variables, 134 were included in the model.

In the final model, a total of 400 variables were included; see Exhibit D5.4.

Respondent Person-Level Poststratification

All 23 proposed one-factor effects were included in the model.

All two-factor effects were present except the State × Race interaction. Out of 137 proposed variables, 129 were included in the model.

For three-factor effects, all levels were present for the Age × Race × Gender, Age × Hispanicity × Gender, State × Age × Gender, State × Race × Gender, and State × Hispanicity × Gender interactions. All the others were affected by variable collapsing or dropping. Out of 227 proposed variables, 176 were included in the model.

In the final model, a total of 328 variables were included; see Exhibit D5.5.

Exhibit D5.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 5: South Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**27**	**27**	
Intercept	1	1	1	All levels present.
State	9	8	8	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**176**	**157**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	9 × 4	24	24	All levels present.
State × Population Density	9 × 4	24	16	Drop (1,1/3), (2,2/3), (4,3); zero. Drop (1,2), (2,1), (8,3); sing.
State × Group Quarter	9 × 3	16	8	Coll. (1,1) & (1,2), then drop; zero/conv. Drop (4,2), (5,2), (7,1), (8,2); zero. Drop (4,1), (8,1); sing.
State × % Black or African American	9 × 3	16	16	All levels present.
State × % Hispanic or Latino	9 × 3	16	13	Drop (4,1), (8,*); zero.
State × % Owner-Occupied	9 × 3	16	16	All levels present.
State × Rent/Housing	9 × 5	32	32	All levels present.
Three-Factor Effects		**256**	**122**	
State × % Owner-Occupied × % Black or African American	9 × 3 × 3	32	19	Drop (1,3,1), (6,3,2), (8,3,1/2), (8,2,1); zero. Drop (5,3,2), (7,3,1/2),(1,3,2), (6,3,1), (8,2,2), (4,3,2); sing. Drop (4,3,1); conv.
State × % Owner-Occupied × % Hispanic or Latino	9 × 3 × 3	32	13	Drop (5,3,1/2), (7,3,1), (1,3,1/2), (1,2,1), (1,2,1), (6,3,1/2), (6,2,1), (8,*,*), (4,2/3,1); zero. Drop (5,2,1), (7,2,1), (2,3,1), (2,2,1); sing.
State × % Owner-Occupied × Rent/Housing	9 × 3 × 5	64	30	Drop (5,3,1), (7,3,1), (2,2,1), (1,3,1/2/3), (1,2,1), (6,3,2/3/4), (8,3,1/3/4),(8,2,3/4), (4,3,3); zero. Drop (5,3,3/4), (5,2,4), (7,3,2/3/4), (7,2,3), (2,3,1), (2,2,2), (1,3,4), (6,3,1), (8,3,2), (8,2,2); sing. Drop (7,2,4), (4,3,1/2), (4,3,4), coll. (4,2,1) & (4,2,2); conv.
State × Rent/Housing × % Black or African American	9 × 3 × 5	64	42	Drop (5,4,1), (7,3,1), (2,1,2), (8,2,1), (8,3,1/2), (8,2,1),(8,4,1/2), (4,1,2); zero. Drop (9,4,1), (5,3,1), (7,4,1), (2,1,1), (2,2,2), (2,4,2), (1,1,2), (1,4,1), (6,4,1), (8,1,1), (8,2,2), (4,4,1); sing.
State × Rent/Housing × % Hispanic or Latino	9 × 3 × 5	64	18	Keep (9,1,2), (9,2,2), (9,3,2), (9,4,1/2), (5,1,1), (5,2,2), (5,3,2),(7,2,2),(7,3,2), (7,4,2), (2,2,2), (2,3,2), (2,4,2), (1,2,2), (1,3,2), (6,2,2), (4,3,2), drop all others; sing./zero.
Total		**459**	**306**	

Exhibit D5.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 5: South Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**22**	**22**	
Intercept	1	1	1	All levels present.
State	9	8	8	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**125**	**124**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	9 × 4	24	24	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race (5 levels)	9 × 5	32	31	Coll. (1,3) & (1,4); conv.
State × Hispanicity	9 × 2	8	8	All levels present.
State × Gender	9 × 2	8	8	All levels present.
Three-Factor Effects		**190**	**187**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	9 × 5 × 3	64	64	All levels present.
State × Age × Hispanicity	9 × 5 × 2	32	31	Coll. (8,3,1) & (8,4,1); conv.
State × Age × Gender	9 × 5 × 2	32	32	All levels present.
State × Race (3 levels) × Hispanicity	9 × 3 × 2	16	14	Coll. (5,3,1) & (5,3,1); (8,2,1) & (8,3,1); conv.
State × Race (3 levels) × Gender	9 × 3 × 2	16	16	All levels present.
State × Hispanicity × Gender	9 × 2 × 2	8	8	All levels present.
Total		**337**	**333**	

Exhibit D5.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 5: South Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**40**	**40**	
Intercept	1	1	1	All levels present.
State	9	8	8	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**237**	**232**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	9 × 4	24	24	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race (5 levels)	9 × 5	32	31	Coll. (4,3) & (4,4); conv.
State × Hispanicity	9 × 2	8	8	All levels present.
State × Gender	9 × 2	8	8	All levels present.
State × % Black or African American	9 × 3	16	16	All levels present.
State × % Hispanic or Latino	9 × 3	16	13	Drop (3,1), (8,1/2); zero.
State × % Owner-Occupied	9 × 3	16	16	All levels present.
State × Rent/Housing	9 × 5	32	31	Drop (8,4); sing.
Three-Factor Effects		**190**	**168**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	6	Coll. (4,2,1) & (4,3,1), repeat for age level 3; conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	9 × 5 × 3	64	61	Coll. (8,4,2) & (8,4,3), repeat for age levels 3 and 2; conv.
State × Age × Hispanicity	9 × 5 × 2	32	24	Drop (8,3/4,1); zero. Drop (8,1/2,1), (6,*,1); conv.
State × Age × Gender	9 × 5 × 2	32	32	All levels present.
State × Race (3 levels) × Hispanicity	9 × 3 × 2	16	7	Coll. (8,2,1) & (8,3,1), then drop; zero/conv. Coll. (9,2,1) & (9,3,1), repeat for all other States; conv.
State × Race (3 levels) × Gender	9 × 3 × 2	16	16	All levels present.
State × Hispanicity × Gender	9 × 2 × 2	8	8	All levels present.
Total		**467**	**440**	

Exhibit D5.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 5: South Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**40**	**40**	
Intercept	1	1	1	All levels present.
State	9	8	8	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**237**	**226**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	9 × 4	24	24	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race (5 levels)	9 × 5	32	25	Coll. (1,3) & (1,4), repeat for all States except FL; conv.
State × Hispanicity	9 × 2	8	8	All levels present.
State × Gender	9 × 2	8	8	All levels present.
State × % Black or African American	9 × 3	16	16	All levels present.
State × % Hispanic or Latino	9 × 3	16	13	Drop (4,1), (8,1/2); zero
State × % Owner-Occupied	9 × 3	16	16	All levels present.
State × Rent/Housing	9 × 5	32	31	Drop (8,4); sing.
Three-Factor Effects		**190**	**134**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (2,2,1) & (2,3,1), repeat for age levels 3 and 4; conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	9 × 5 × 3	64	48	Coll. (1,3,2) & (1,3,3), (1,4,2) & (1,4,3), repeat for State NC; conv. Coll. (6,1,2) & (6,1,3), repeat for all age levels and for States VA and WV; conv.
State × Age × Hispanicity	9 × 5 × 2	32	16	Drop (5,4,1); sing. Drop (8,2/3/4,1); zero. Drop rest for States NC, SC, VA; conv.
State × Age × Gender	9 × 5 × 2	32	32	All levels present.
State × Race (3 levels) × Hispanicity	9 × 3 × 2	16	2	Drop (8,2,1); zero. Coll. (1,2,1) & (1,3,1), repeat for States MD, NC, SC, and VA, then drop the collapsed variables and (8,1,1); conv.
State × Race (3 levels) × Gender	9 × 3 × 2	16	9	Coll. (1,2,1) & (1,3,1), repeat for all States except FL; conv.
State × Hispanicity × Gender	9 × 2 × 2	8	8	All levels present.
Total		**467**	**400**	

Exhibit D5.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 5: South Atlantic

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**23**	**23**	
Intercept	1	1	1	All levels present.
State	9	8	8	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**137**	**129**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	9 × 4	24	24	All levels present.
State × Age	9 × 6	40	40	All levels present.
State × Race (5 levels)	9 × 5	32	24	Coll. (2,3) & (2,4), (4,3) & (4,4), (7,3) & (7,4); conv. Repeat for the rest of the States; conv.
State × Hispanicity	9 × 2	8	8	All levels present.
State × Gender	9 × 2	8	8	All levels present.
Three-Factor Effects		**227**	**176**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	5	Coll. (1,2,1) & (1,3,1), repeat for all age levels; conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	9 × 6 × 3	80	63	Coll. (2,5,2) & (2.5.3), (6,5,2) & (6,5,3), (8,5,2) & (8,5,3); sing. Coll. (1,1,2) & (1,1,3), repeat for all age levels for States DE, VA, and WV; conv.
State × Age × Hispanicity	9 × 6 × 2	40	21	Drop (2,5,1), (1,5,1), (5,4,1), (8,2,1); sing; Drop (5,5,1), (8,3/4/5,1); zero. Drop all for States SC, VA, and WV; conv.
State × Age × Gender	9 × 6 × 2	40	40	All levels present.
State × Race (3 levels) × Hispanicity	9 × 3 × 2	16	7	Coll. (1,2,1) & (1,3,1), repeat for all States, drop all for State SC; conv.
State × Race (3 levels) × Gender	9 × 3 × 2	16	16	All levels present.
State × Hispanicity × Gender	9 × 2 × 2	8	8	All levels present.
Total		**387**	**328**	

Appendix D6: Model Group 6: East South Central

(Alabama, Kentucky, Mississippi, and Tennessee)

Table D.6a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 6: East South Central)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	2.58	3.25	0.09	1.09241	204	(1.03, 1.45)	(1.04, 1.45)
	1.20	1.54	0.10	1.09735	121	(1.00, 2.45)	(1.00, 2.44)
						(1.00, 1.20)	(1.00, 1.01)
res.sdu.ps	1.20	1.54	0.10	1.09736	162	(0.53, 1.10)	(0.53, 1.10)
	1.85	3.39	0.51	1.14655	150	(0.20, 4.31)	(0.20, 4.28)
						(0.95, 1.05)	(1.05, 1.05)
sel.per.ps	3.33	6.42	1.42	2.25471	242	(0.20, 1.57)	(0.20, 1.57)
	1.01	1.67	0.32	2.32228	212	(0.20, 4.90)	(0.20, 4.90)
						(0.70, 1.77)	(0.70, 1.64)
res.per.nr	1.15	1.87	0.38	2.37375	242	(1.00, 2.60)	(1.00, 2.60)
	1.18	2.31	0.54	2.56870	187	(1.00, 4.80)	(1.00, 4.79)
						(1.20, 1.70)	(1.70, 1.70)
res.per.ps	1.26	2.70	0.62	2.56870	187	(0.20, 1.20)	(0.20, 1.20)
	1.09	2.79	0.37	2.69088	145	(0.20, 4.06)	(0.20, 4.02)
						(0.95, 1.05)	(0.95, 0.95)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.6b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 6: East South Central)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	128	0.97	255	0.20	97	1.01	103	0.13	21	0.34	21	0.10	5
1%	447	1.00	464	0.36	234	1.01	267	0.23	195	1.00	260	0.21	115
5%	455	1.01	488	0.75	480	1.01	585	0.61	487	1.00	524	0.71	503
10%	504	1.02	528	0.85	555	1.01	708	0.72	640	1.02	720	0.85	699
25%	627	1.05	659	0.96	669	1.11	974	0.87	930	1.07	1,076	0.96	1,080
Median	672	1.08	761	1.06	824	1.43	1,447	1.00	1,495	1.17	1,741	1.01	1,743
75%	859	1.13	970	1.17	1,089	5.67	4,836	1.14	4,822	1.32	5,687	1.05	5,389
90%	1,172	1.15	1,261	1.31	1,337	10.37	8,637	1.34	8,766	1.52	11,136	1.18	10,977
95%	1,208	1.18	1,319	1.45	1,472	11.37	10,724	1.46	11,575	1.74	15,960	1.25	15,579
99%	1,233	1.27	1,415	1.91	1,966	12.99	15,663	2.16	18,208	2.36	24,303	1.77	26,519
Maximum	2,421	2.44	2,381	4.28	3,727	18.93	44,286	4.90	28,573	4.79	45,397	4.02	42,042
n	8,961	8,223	8,223	8,222	8,222	4,534	4,534	4,534	4,534	3,656	3,656	3,656	3,656
Max/Mean	3.15	-	2.85	-	4.15	-	13.35	-	8.41	-	10.78	-	9.98

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 6 Overview

Dwelling Unit Nonresponse

All 22 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Hispanic or Latino, State × Population Density, State × Group Quarter, State × percent Hispanic or Latino, and State × percent Owner-Occupied interactions. Out of 86 proposed variables, 68 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 96 proposed variables, 31 were included in the model.

In the final model, a total of 121 variables were included; see Exhibit D6.1.

Dwelling Unit Poststratification

All 17 proposed one-factor effects were included in the model.

All 60 proposed two-factor effects were included in the model.

For the three-factor effects, variable collapsing or dropping was present in the State × Age × Race, State × Age × Hispanicity, State × Race × Hispanicity, and State × Race × Gender interactions. Out of 85 proposed variables, 73 were included in the model.

In the final model, a total of 150 variables were included; see Exhibit D6.2.

Selected Person-Level Poststratification

All 35 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Hispanic or Latino, and State × percent Hispanic or Latino interactions. Out of 122 proposed variables, 112 were included in the model.

For the three-factor effects, all levels were present for the Age × Race × Gender, Age × Hispanicity × Gender, Race × Hispanicity × Gender, State × Age × Gender, State × Race × Gender, and State × Hispanicity × Gender interactions. Variable collapsing or dropping was present in all other interactions. Out of 85 proposed variables, 65 were included in the model.

In the final model, a total of 212 variables were included; see Exhibit D6.3.

Respondent Person-Level Nonresponse

All 35 proposed one-factor effects were included in the model.

For the two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Hispanic Latino, State × Race, State × percent Hispanic or Latino, and State × percent Owner-Occupied interactions. Out of 122 proposed variables, 105 were included in the model.

For the three-factor effects, all levels were present for the State × Age × Gender and State × Race × Gender interactions. Variable collapsing or dropping was present in all other interactions. Out of 85 proposed variables, 47 were included in the model.

In the final model, a total of 187 variables were included; see Exhibit D6.4.

Respondent Person-Level Poststratification

All 18 proposed one-factor effects were included in the model.

All 67 proposed two-factor effects were included in the model.

Variable collapsing or dropping was present in all three-factor effects except the Age × Race × Gender, State × Age × Gender, and State × Race × Gender interactions. Out of 102 proposed variables, 60 were included in the model.

In the final model, a total of 145 variables were included; see Exhibit D6.5.

Exhibit D6.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 6: East South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**22**	**22**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**86**	**68**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	3	Drop (3,1); zero.
% Owner-Occupied × Rent/Housing	3 × 5	8	7	Drop (3,2); zero.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	4	Drop (1,1), (4,1); zero. Drop (2/3,1); sing.
State × Quarter	4 × 4	9	9	All levels present.
State × Population Density	4 × 4	9	8	Drop (1,3); zero.
State × Group Quarter	4 × 3	6	0	Drop all; zero/sing.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	2	Drop (1,1), (3,2); sing. Drop (2,1), (3,1); zero.
State × % Owner-Occupied	4 × 3	6	5	Coll. (3,2) & (3,3); conv.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**96**	**31**	
State × % Owner-Occupied × % Black or African American	4 × 3 × 3	12	7	Coll. (3,2,1) & (3,3,1), (3,2,2) & (3,3,2); hier. Drop (1,3,1); zero. Drop (1,3,2), (2,3,2); sing.
State × % Owner-Occupied × % Hispanic or Latino	4 × 3 × 3	12	1	Coll. (3,2,1) & (3,3,1), (3,2,2) & (3,3,2), then drop; hier./zero. Keep (1,3,2), drop all others; zero/sing.
State × % Owner-Occupied × Rent/Housing	4 × 3 × 5	24	9	Coll. (3,2,1) & (3,3,1), (3,2,2) & (3,3,2), (3,2,3) & (3,3,3), (3,2,4) & (3,3,4); hier. Drop (3,*,2); zero. Drop (3,*,3/4); sing. Keep (1,3,3), (1,2,*), (2,2,2/3/4), drop all others; sing./zero.
State × Rent/Housing × % Black or African American	4 × 3 × 5	24	13	Drop (2,3,1), (2,4,1); zero. Drop (1,1,2), (1,4,1), (2,2,1), (3,4,1); sing. Coll. (3,3,1) & (3,3,2), drop (3,1,1/2), (3,2,2), (3,4,2); conv.
State × Rent/Housing × % Hispanic or Latino	4 × 3 × 5	24	1	Keep (2,1,2), drop all others; zero/sing.
Total		**204**	**121**	

Exhibit D6.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 6: East South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**17**	**17**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**60**	**60**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	12	All levels present.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
Three-Factor Effects		**85**	**73**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	4 × 5 × 3	24	20	Coll. (3,1,2) & (3,1,3), repeat for all age levels; conv.
State × Age × Hispanicity	4 × 5 × 2	12	11	Drop (3,4,1); conv.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	0	Drop all; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	5	Coll. (3,2,1) & (3,3,1); conv.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**162**	**150**	

Exhibit D6.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 6: East South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**35**	**35**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**122**	**112**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic	3 × 3	4	3	Drop (3,1); zero.
% Owner-Occupied × Rent/Housing	3 × 5	8	7	Drop (3,2); zero.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	4	Drop (1,1), (4,1); zero. Drop (2,1), (3,1); sing.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	12	All levels present.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	2	Drop (2,1), (3,1); zero. Coll. (1,1) & (1,2), drop (3,2); sing.
State × % Owner-Occupied	4 × 3	6	6	All levels present.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**85**	**65**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (3,2,1) & (3,3,1); zero. Drop (4,2/3,1); sing./conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	4 × 5 × 3	24	14	Coll. (1,1,2) & (1,1,3), repeat for all age groups, for States (1) & (3); conv. Coll. (2,3,2) & (2,3,3), (2,4,2) & (2,4,3); conv.
State × Age × Hispanicity	4 × 5 × 2	12	11	Drop (1,4,1); conv.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	0	Drop all levels; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**242**	**212**	

Exhibit D6.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 6: East South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**35**	**35**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**122**	**105**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	3	Drop (3,1); zero.
% Owner-Occupied × Rent/Housing	3 × 5	8	7	Drop (3,2); zero.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	4	Drop (1,1), (4,1); zero. Drop (2,1), (3,1); sing..
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	8	Coll. (1,3) & (1,4) & (1,5), (2,3) & (2,4), (3,3) & (3,4); conv.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	1	Keep (2,2). Drop all others; zero, sing., conv.
State × % Owner-Occupied	4 × 3	6	4	Coll. (1,3) & (1,2), (3,3) & (3,2); conv.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**85**	**47**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	2	Keep (1,2/3,1), (2,2/3,1). Drop all others; zero/sing./conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	4	Coll. (1,2,1) & (1,3,1), repeat for all age levels; conv.
Age × Hispanicity × Gender	5 × 2 × 2	4	3	Drop (4,1,1); conv.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	0	Drop all; conv.
State × Age × Race (3 levels)	4 × 5 × 3	24	12	Coll. (1,1,2) & (1,1,3), repeat for all age levels and all States; conv.
State × Age × Hispanicity	4 × 5 × 2	12	8	Drop (1,4,1), (2,4,1), (3,3,1), (3,4,1); conv.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	0	Drop all; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	0	Drop all; conv.
Total		**242**	**187**	

Exhibit D6.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 6: East South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**18**	**18**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**67**	**67**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 6	15	15	All levels present.
State × Race (5 levels)	4 × 5	12	12	All levels present.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
Three-Factor Effects		**102**	**60**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	3	Keep (1,2/3,1), (2,2/3,1), (3/4,2/3,1). Drop all others; zero/sing./conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	4	Drop (5,1,1); sing.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	4 × 6 × 3	30	15	Coll. (1,1,2) & (1,1,3), repeat for all States and all age levels; conv.
State × Age × Hispanicity	4 × 6 × 2	15	6	Keep (1,1,1), (1,2,1), repeat for all States, drop all others; zero/sing./conv.
State × Age × Gender	4 × 6 × 2	15	15	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	0	Drop all; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	0	Drop all; conv.
Total		**187**	**145**	

Appendix D7: Model Group 7: West South Central

(Arkansas, Louisiana, Oklahoma, and Texas)

Table D.7a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 7: West South Central)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	1.74	1.88	0.03	1.14549	204	(1.07, 1.20)	(1.07, 1.19)
	1.55	2.32	0.21	1.16879	135	(1.00, 2.90)	(1.00, 2.89)
						(1.10, 1.18)	(1.10, 1.18)
res.sdu.ps	1.55	2.32	0.21	1.16879	162	(0.80, 1.10)	(0.80, 1.10)
	1.19	2.37	0.66	1.20137	160	(0.25, 4.97)	(0.26, 4.96)
						(0.90, 1.23)	(0.90, 1.23)
sel.per.ps	3.54	8.20	2.34	2.19865	242	(0.22, 1.90)	(0.22, 1.90)
	1.89	3.70	0.75	2.14432	236	(0.21, 4.08)	(0.21, 4.07)
						(0.90, 2.59)	(0.90, 2.59)
res.per.nr	2.05	3.97	0.88	2.18841	242	(1.00, 2.49)	(1.00, 2.47)
	1.65	3.34	0.58	2.47148	209	(1.00, 4.76)	(1.00, 4.72)
						(1.30, 5.00)	(1.30, 1.30)
res.per.ps	1.76	3.56	0.61	2.47148	187	(0.20, 1.20)	(0.20, 1.20)
	0.84	1.20	0.17	2.55967	168	(0.20, 3.20)	(0.20, 3.19)
						(0.30, 5.00)	(N/A, N/A)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.7b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 7: West South Central)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	178	0.96	210	0.26	76	1.01	83	0.08	18	0.35	39	0.10	19
1%	383	1.00	392	0.52	340	1.01	401	0.43	311	1.00	381	0.20	226
5%	493	1.03	510	0.78	462	1.01	547	0.66	510	1.02	614	0.20	407
10%	503	1.04	538	0.87	530	1.01	710	0.75	688	1.05	811	0.69	625
25%	580	1.06	641	0.98	701	1.16	1,171	0.88	1,150	1.11	1,331	0.98	1,292
Median	1,016	1.09	1,095	1.07	1,080	1.39	1,758	1.01	1,799	1.21	2,128	1.05	2,214
75%	1,101	1.15	1,223	1.18	1,327	5.32	5,473	1.13	5,493	1.37	6,435	1.14	6,079
90%	1,141	1.21	1,311	1.30	1,509	9.61	8,824	1.28	9,287	1.59	12,239	1.23	12,353
95%	1,166	1.25	1,411	1.41	1,659	10.03	12,761	1.39	12,746	1.77	17,475	1.28	17,185
99%	2,149	1.46	2,386	1.86	2,720	11.64	16,663	1.88	16,718	2.55	25,941	1.69	27,886
Maximum	3,064	2.89	3,862	4.96	6,541	27.54	64,077	5.08	38,094	4.72	58,131	3.19	45,485
n	14,468	13,032	13,032	13,032	13,032	8,113	8,113	8,113	8,113	6,347	6,347	6,347	6,347
Max/Mean	3.46	-	3.93	-	6.14	-	17.33	-	10.26	-	12.25	-	9.58

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 7 Overview

Dwelling Unit Nonresponse

Out of 22 proposed one-factor effects, 21 were included in the model. Variable collapsing was present in the Group Quarter main effect.

For two-factor effects, variable collapsing and dropping was present in the percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Black or African American, State × Population Density, State × Group Quarter, and State × percent Hispanic or Latino interactions. Out of 86 proposed variables, 78 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 96 proposed variables, 36 were included in the model.

In the final model, a total of 135 variables were included; see Exhibit D7.1.

Dwelling Unit Poststratification

All 17 proposed one-factor effects were included in the model.

All 60 proposed two-factor effects were included in the model.

For three-factor effects, variable collapsing and dropping were present in the Age × Race × Hispanicity interaction. Out of 85 proposed variables, 83 were included in the model.

In the final model, a total of 160 variables were included; see Exhibit D7.2.

Selected Person-Level Poststratification

All 35 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing was present in the Rent/Housing × percent Black or African American and State × Race interactions. Out of 122 proposed variables, 118 were included in the model.

For three-factor effects, variable collapsing and dropping were present in the State × Race × Hispanicity interaction. Out of 85 proposed variables, 83 were included in the model.

In the final model, a total of 236 variables were included; see Exhibit D7.3.

Respondent Person-Level Nonresponse

All 35 proposed one-factor effects were included in the model.

For two-factor effects, variable dropping was present in the Rent/Housing × percent Black or African American and State × percent Hispanic or Latino interactions. Out of 122 proposed variables, 120 were included in the model.

Variable collapsing or dropping was present in the Race × Hispanicity × Gender, State × Age × Race, State × Age × Hispanicity, and State × Race × Hispanicity interactions. Out of 85 proposed variables, 54 were included in the model.

In the final model, a total of 209 variables were included; see Exhibit D7.4.

Respondent Person-Level Poststratification

All 18 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing was present in the State × Race interaction. Out of 67 proposed variables, 64 were included in the model.

For three-factor effects, variable collapsing and dropping were present in Age × Race × Hispanicity, Race × Hispanicity × Gender, State × Age × Race, State × Age × Hispanicity, and State × Race × Hispanicity interactions. Out of 102 proposed variables, 86 were included in the model.

In the final model, a total of 168 variables were included; see Exhibit D7.5.

Exhibit D7.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 7: West South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**22**	**21**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	1	Coll. (1) & (2); conv.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**86**	**78**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	7	Coll. (3,1) & (3,2); conv.
Rent/Housing × % Black or African American	3 × 5	8	7	Drop (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Population Density	4 × 4	9	8	Drop (2,3); sing.
State × Group Quarter	4 × 3	6	2	Coll. (4,1) & (4,2), repeat for all States; hier. Drop (3,*); sing.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	5	Drop (2,1); sing.
State × % Owner-Occupied	4 × 3	6	6	All levels present.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**96**	**36**	
State × % Owner-Occupied × % Black or African American	4 × 3 × 3	12	5	Keep (2,2,1/2), (3,2,2), (4,2,1/2), drop all others; zero/sing.
State × % Owner-Occupied × % Hispanic or Latino	4 × 3 × 3	12	3	Keep (2,2,2), (4,2,1/2), drop all others; zero/sing./conv.
State × % Owner-Occupied × Rent/Housing	4 × 3 × 5	24	6	Keep (2,2,1/2/3), (3,2,1/2/3), drop all others; zero/sing./conv.
State × Rent/Housing × % Black or African American	4 × 3 × 5	24	17	Drop (2,4,1), (3.1.1), (3,4,2), (4,4,1); sing. Drop (3,3,1), (3,2,1), (3,4,1); zero.
State × Rent/Housing × % Hispanic or Latino	4 × 3 × 5	24	5	Keep (3,2,2), (3,3,2), (4,2,2), (4,1,2), (4,3,2), drop all others; sing./zero.
Total		**204**	**135**	

Exhibit D7.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 7: West South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**17**	**17**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**60**	**60**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	12	All levels present.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
Three-Factor Effects		**85**	**83**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	6	Coll. (3,2,1) & (3,3,1), (4,2,1) & (4,3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	4 × 5 × 3	24	24	All levels present.
State × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**162**	**160**	

Exhibit D7.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 7: West South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**35**	**35**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**122**	**118**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	7	Drop (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	10	Coll. (2,3) & (2,4), (3,3) & (3,4); conv.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	5	Coll. (2,1); zero.
State × % Owner-Occupied	4 × 3	6	6	All levels present.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**85**	**83**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	4 × 5 × 3	24	24	All levels present.
State × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	4	Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**242**	**236**	

Exhibit D7.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 7: West South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**35**	**35**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**122**	**120**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	7	Drop (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 5	12	12	All levels present.
State × Race (5 levels)	4 × 5	12	12	All levels present.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
State × % Black or African American	4 × 3	6	6	All levels present.
State × % Hispanic or Latino	4 × 3	6	5	Drop (2,1); zero.
State × % Owner-Occupied	4 × 3	6	6	All levels present.
State × Rent/Housing	4 × 5	12	12	All levels present.
Three-Factor Effects		**85**	**54**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	4 × 5 × 3	24	12	Coll. (2,1,2) & (2,1,3), repeat for all age levels; conv. Coll. (3,1,2) & (3,1,3), repeat for all age levels, then drop; conv.
State × Age × Hispanicity	4 × 5 × 2	12	0	Drop (3,4,1); sing. Drop the rest; conv.
State × Age × Gender	4 × 5 × 2	12	12	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	0	Coll. (2,2,1) & (2,3,1), repeat for all States, then drop; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**242**	**209**	

Exhibit D7.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 7: West South Central

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**18**	**18**	
Intercept	1	1	1	All levels present.
State	4	3	3	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**67**	**64**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	4 × 4	9	9	All levels present.
State × Age	4 × 6	15	15	All levels present.
State × Race (5 levels)	4 × 5	12	9	Coll. (4,3) & (4,4), repeat for all States; conv.
State × Hispanicity	4 × 2	3	3	All levels present.
State × Gender	4 × 2	3	3	All levels present.
Three-Factor Effects		**102**	**86**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	3	Coll. (1,1,2) & (1,1,3), repeat for all age levels, then drop (4,*,1), (5,*,1); conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	4 × 6 × 3	30	29	Coll. (2,5,2) & (2,5,3); sing..
State × Age × Hispanicity	4 × 6 × 2	15	11	Drop (3,5,1); sing. Drop (2,3/4/5,1); conv.
State × Age × Gender	4 × 6 × 2	15	15	All levels present.
State × Race (3 levels) × Hispanicity	4 × 3 × 2	6	3	Coll. (4,2,1) & (4,3,1), repeat for all States; conv.
State × Race (3 levels) × Gender	4 × 3 × 2	6	6	All levels present.
State × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
Total		**187**	**168**	

Appendix D8: Model Group 8: Mountain

(Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming)

Table D.8a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 8: Mountain)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	4.48	10.30	2.70	1.76798	408	(1.02, 2.30)	(1.02, 2.30)
	3.68	7.47	1.07	1.75070	196	(1.00, 2.59)	(1.00, 2.59)
						(1.00, 1.70)	(1.01, 1.68)
res.sdu.ps	3.68	7.48	1.07	1.75075	302	(0.29, 1.30)	(0.29, 1.30)
	2.20	3.64	0.76	1.73678	290	(0.20, 5.00)	(0.20, 5.00)
						(0.90, 3.83)	(0.90, 3.83)
sel.per.ps	3.52	7.50	1.66	3.37913	422	(0.20, 3.00)	(0.20, 3.00)
	1.69	3.53	0.79	3.46434	384	(0.20, 4.87)	(0.21, 4.86)
						(0.90, 3.14)	(0.90, 3.14)
res.per.nr	1.84	3.86	0.92	3.47054	422	(1.00, 3.00)	(1.00, 3.00)
	1.51	3.85	0.78	3.81372	355	(1.00, 5.00)	(1.00, 4.99)
						(1.20, 1.37)	(1.20, 1.36)
res.per.ps	1.51	3.67	0.84	3.81372	347	(0.20, 1.80)	(0.20, 1.77)
	0.95	2.35	0.31	4.02892	305	(0.20, 4.88)	(0.20, 4.86)
						(0.90, 1.11)	(0.90, 1.11)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.8b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 8: Mountain)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	30	0.56	73	0.14	20	1.01	29	0.07	8	0.37	19	0.11	12
1%	70	0.93	76	0.44	80	1.01	98	0.31	71	0.95	75	0.21	59
5%	83	1.00	91	0.80	98	1.01	146	0.59	141	1.00	165	0.39	141
10%	91	1.02	98	0.92	111	1.01	189	0.71	189	1.02	225	0.74	207
25%	139	1.05	150	1.06	178	1.17	418	0.86	400	1.08	456	0.96	417
Median	295	1.08	317	1.19	405	1.48	951	0.99	907	1.17	1,052	1.02	1,029
75%	561	1.12	612	1.35	748	5.59	2,132	1.15	2,185	1.32	2,630	1.08	2,533
90%	763	1.18	871	1.62	1,112	9.41	5,395	1.33	5,261	1.55	6,478	1.26	6,298
95%	807	1.25	1,017	1.81	1,356	12.00	7,679	1.47	7,919	1.77	10,203	1.45	10,292
99%	1,440	1.71	1,627	2.78	2,062	13.65	15,911	1.96	16,115	2.37	20,186	1.73	20,727
Maximum	3,062	4.04	3,189	6.33	4,992	22.77	48,558	7.54	52,100	4.99	55,524	4.86	69,374
n	17,185	15,671	15,671	15,670	15,670	9,054	9,054	9,054	9,054	7,282	7,282	7,282	7,282
Max/Mean	7.66	-	7.28	-	9.29	-	23.86	-	25.60	-	21.95	-	27.42

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 8 Overview

Dwelling Unit Nonresponse

For one-factor effects, the Group Quarter main effect was dropped due to the convergent problem. Out of 26 proposed variables, 24 were included in the model.

For two-factor effects, all levels were present in the percent Owner-Occupied × percent Hispanic or Latino, percent Owner-Occupied × Rent/Housing, Rent/Housing × percent Hispanic or Latino, State × Group Quarter, and State × Rent/Housing interactions. All the others were affected by variable collapsing or dropping. Out of 158 proposed variables, 117 were included in the model.

All three-factor effects were affected by variable collapsing and dropping. Out of 224 proposed variables, 55 were included in the model.

In the final model, a total of 196 variables were included; see Exhibit D8.1.

Dwelling Unit Poststratification

All 21 proposed one-factor effects were included in the model.

All 112 proposed two-factor effects were included in the model.

For three-factor effects, variable collapsing was present in the Age × Race ×Hispanicity, State × Age × Race, State × Race × Hispanicity, and State × Race × Hispanicity interactions. Out of 169 proposed variables, 157 were included in the model.

In the final model, a total of 290 variables were included; see Exhibit D8.2.

Selected Person-Level Poststratification

All 39 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Black or African American, Rent/Housing × percent Black or African American, State × percent Black or African American, State × percent Hispanic or Latino, and State × percent Owner-Occupied interactions. Out of 214 proposed variables, 195 were included in the model.

For three-factor effects, variable collapsing was present in the Age × Race × Hispanicity, State × Age × Race, State × Age × Hispanicity, State × Race × Hispanicity, and State × Race × Gender interactions. Out of 169 proposed variables, 150 were included in the model.

In the final model, a total of 384 variables were included; see Exhibit D8.3.

Respondent Person-Level Nonresponse

All 39 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the percent Owner-Occupied × percent Black or African American, percent Owner-Occupied × Rent/Housing, State × Race, State × percent Black or African American, State × percent Hispanic or Latino, and State × percent Owner-Occupied interactions. Out of 214 proposed variables, 192 were included in the model.

For three-factor effects, all levels were present in the Age × Race × Gender, Age × Hispanicity × Gender, Race × Hispanicity × Gender, State × Age × Hispanicity, and State × Age × Gender interactions. All the others were affected by variable collapsing or dropping. Out of 169 proposed variables, 124 were included in the model.

In the final model, a total of 355 variables were included; see Exhibit D8.4.

Respondent Person-Level Poststratification

All 22 proposed one-factor effects were included in the model.

All 123 proposed two-factor effects were included in the model.

Variable collapsing or dropping was present in all three-factor effects except the Age × Hispanicity × Gender, State × Age × Gender, State × Race × Gender, and State × Hispanicity × Gender interactions. Out of 202 proposed variables, 160 were included in the model.

In the final model, a total of 305 variables were included; see Exhibit D8.5.

Exhibit D8.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 8: Mountain

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**26**	**24**	
Intercept	1	1	1	All levels present.
State	8	7	7	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	0	Drop all; conv.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**158**	**117**	
% Owner-Occupied × % Black or African American	3 × 3	4	2	Drop (3,1); sing. Drop (2,1); zero.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	4	Drop (1,1), (1,2) & (1,3); zero. Drop (1,4); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	8 × 4	21	21	All levels present.
State × Population Density	8 × 4	21	15	Drop (2,1), (3,1), (5,1), (6,1), (7,1) & (7,3); zero.
State × Group Quarter	8 × 3	14	0	Drop (*,*); hier.
State × % Black or African American	8 × 3	14	3	Keep (1,2), (4,2), (5,2); Drop remainder; sing./zero.
State × % Hispanic or Latino	8 × 3	14	12	Drop (3,1), (7,1); zero.
State × % Owner-Occupied	8 × 3	14	12	Drop (2,2) & (2,3), (3,2) & (3,3); conv.
State × Rent/Housing	8 × 5	28	28	All levels present.
Three-Factor Effects		**224**	**55**	
State × % Owner-Occupied × % Black or African American	8 × 3 × 3	28	3	Keep (1,2,2), (4,2,2), (7,2,2). Drop remainder; zero, sing.
State × % Owner-Occupied × % Hispanic or Latino	8 × 3 × 3	28	10	Keep (1,2,*), (2,2+3,2), (4,2,*), (4,3,1), (5,2,1), (6,*,2), (7,2,2). Coll. (2,2,2) &(2,3,2); hier. Drop remainder; zero/sing./conv.
State × % Owner-Occupied × Rent/Housing	8 × 3 × 5	56	20	Keep (1,3,4), (1,2,3), (1,2,4), (4,2,2), (4,*,3), (4,2,4), (5,3,2), (5,2,*), (6,2,2), (6,2,3), (6,2,4), (7,2,1), (7,2,2), (7,2,3); conv. Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); hier. Drop remainder; zero/sing./conv.
State × Rent/Housing × % Black or African American	8 × 3 × 5	56	2	Keep (1,3,2), (1,4,2). Drop remainder; zero/sing.
State × Rent/Housing × % Hispanic or Latino	8 × 3 × 5	56	20	Keep (1,1,1), (1,2,1), (1,3,1), (1,3,2), (1,4,2), (2,1,2), (2,2,2), (4,1,2), (4,2,1), (4,3,1), (4,4,2), (5,1,1), (5,1,2), (5,2,1), (6,1,2), (6,2,2), (6,3,2), (6,4,2) & (7,1,2), (7,2,2). Drop remainder; zero/sing./conv.
Total		**408**	**196**	

Exhibit D8.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 8: Mountain

Variables	Levels	Proposed	Final	Comment
One-Factor Effects		**21**	**21**	
Intercept	1	1	1	All levels present.
State	8	7	7	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**112**	**112**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	8 × 4	21	21	All levels present.
State × Age	8 × 5	28	28	All levels present.
State × Race (5 levels)	8 × 5	28	28	All levels present.
State × Hispanicity	8 × 2	7	7	All levels present.
State × Gender	8 × 2	7	7	All levels present.
Three-Factor Effects		**169**	**157**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (2,2,1) & (2,3,1), repeat for all age levels; conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	8 × 5 × 3	56	51	Coll. (3,1,2) & (3,1,3); zero. Coll. (3,2,2) & (3,2,3), (7,1,2) & (7,1,3), (7,2,2) & (7,2,3), (7,3,2) & (7,3,3); conv.
State × Age × Hispanicity	8 × 5 × 2	28	28	All levels present.
State × Age × Gender	8 × 5 × 2	28	28	All levels present.
State × Race (3 levels) × Hispanicity	8 × 3 × 2	14	11	Coll. (6,2,1) & (6,3,1); conv. Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); zero.
State × Race (3 levels) × Gender	8 × 3 × 2	14	14	All levels present.
State × Hispanicity × Gender	8 × 2 × 2	7	7	All levels present.
Total		**302**	**290**	

Exhibit D8.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 8: Mountain

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**39**	**39**	
Intercept	1	1	1	All levels present.
State	8	7	7	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**214**	**195**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	2	Drop (3,1); sing. Drop (2,1); zero.
% Owner-Occupied × % Hispanic	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	4	Drop (1,1),(2,1),(3,1); zero. Drop (4,1); sing..
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	8 × 4	21	21	All levels present.
State × Age	8 × 5	28	28	All levels present.
State × Race (5 levels)	8 × 5	28	28	All levels present.
State × Hispanicity	8 × 2	7	7	All levels present.
State × Gender	8 × 2	7	7	All levels present.
State × % Black or African American	8 × 3	14	4	Drop (1,1), (2,*), (3,*), (5,1), (6,1), (7,*); zero. Drop (4,1); sing.
State × % Hispanic or Latino	8 × 3	14	12	Drop (3,1), (7,1); zero.
State × % Owner-Occupied	8 × 3	14	13	Drop (3,3); zero.
State × Rent/Housing	8 × 5	28	28	All levels present.
Three-Factor Effects		**169**	**150**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	7	Coll. (4,2,1), (4,3,1); sing.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	8 × 5 × 3	56	46	Coll. (2,3,2) & (2,3,3), (3,1,2) & (3,1,3); zero. Coll. (6,4,2) & (6,4,3); sing. Coll. (2,1,2) & (2,1,3), (2,2,2) & (2,2,3), (2,4,2) & (2,4,3), (3,2,2) & (3,2,3), (5,3,2) & (5,3,3), (7,3,2) & (7,3,3) (7,4,2) & (7,4,3); conv.
State × Age × Hispanicity	8 × 5 × 2	28	27	Drop (3,4,1); conv.
State × Age × Gender	8 × 5 × 2	28	28	All levels present.
State × Race (3 levels) × Hispanicity	8 × 3 × 2	14	9	Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); zero. Coll. (6,2,1) & (6,3,1); sing. Coll. (5,2,1) & (5,3,1), (7,2,1) & (7,3,1); conv.
State × Race (3 levels) × Gender	8 × 3 × 2	14	12	Coll. (3,2,1) & (3,3,1), (5,2,1) & (5,3,1); conv.
State × Hispanicity × Gender	8 × 2 × 2	7	7	All levels present.
Total		**422**	**384**	

Exhibit D8.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 8: Mountain

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**39**	**39**	
Intercept	1	1	1	All levels present.
State	8	7	7	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**214**	**192**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	2	Drop (2,1); sing. Drop (3,1); sing.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	4	Drop (4,1); sing. Drop (1,1), (1,2) & (1,3); zero.
Rent/Housing × % Black or African American	3 × 5	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	8 × 4	21	21	All levels present.
State × Age	8 × 5	28	28	All levels present.
State × Race (5 levels)	8 × 5	28	27	Coll. (3,2) & (3,5); conv.
State × Hispanicity	8 × 2	7	7	All levels present.
State × Gender	8 × 2	7	7	All levels present.
State × % Black or African American	8 × 3	14	3	Drop (1,1), (2,*), (3,*), (5,1), (6,1) & (7,*); zero. Drop (4,1); sing. Drop (6,2); conv.
State × % Hispanic or Latino	8 × 3	14	12	Drop (3,1) & (7,1); zero.
State × % Owner-Occupied	8 × 3	14	12	Coll. (1,2) & (1,3); conv. Drop (3,3); zero.
State × Rent/Housing	8 × 5	28	28	All levels present.
Three-Factor Effects		**169**	**124**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	7	Drop (4,2,1); sing.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	8 × 5 × 3	56	22	Coll. (1,1,2), (1,1,3), (1,2,2) & (1,2,3), (1,3,2) & (1,3,3), (1,4,2) & (1,4,3). (2,1,2), (2,1,3), (2,2,2) & (2,2,3), (4,1,2), (4,1,3), (4,2,2) & (4,2,3), (5,1,2) & (5,1,3), (5,2,2) & (5,2,3), (5,3,2) & (5,3,3), (5,4,2) & (5,4,3), (6,1,2), (6,1,3), (6,2,2) & (6,2,3), (6,4,2) & (6,3,3), (7,1,2) & (7,1,3), (7,2,2) & (7,2,3), (7,3,2) & (7,3,3), (7,4,2) & (7,4,3); conv. Coll. (3,2,2) & (3,2,3), (3,3,2) & (3,3,3), (3,4,2) & (3,4,3); hier. Drop (2,3,2), (3,1,2); zero. Drop (2,4,2), (6,4,2); sing. Drop (2,4,3), (4,3,*), (4,4,*), (6,4,*); conv.
State × Age × Hispanicity	8 × 5 × 2	28	28	All levels present.
State × Age × Gender	8 × 5 × 2	28	28	All levels present.
State × Race (3 levels) × Hispanicity	8 × 3 × 2	14	6	Drop (1,*,1), (5,*,1); conv. Drop (2,2,1); zero. Coll. (3,2,1) & (3,3,1); hier. Coll. (4,2,1) & (4,3,1); conv. Drop (6,2,1); sing.
State × Race (3 levels) × Gender	8 × 3 × 2	14	12	Coll. (1,2,1) & (1,3,1); conv. Coll. (3,2,1) & (3,3,1); hier.
State × Hispanicity × Gender	8 × 2 × 2	7	7	All levels present.
Total		**422**	**355**	

Exhibit D8.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 8: Mountain

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**22**	**22**	
Intercept	1	1	1	All levels present.
State	8	7	7	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**123**	**123**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	8 × 4	21	21	All levels present.
State × Age	8 × 6	35	35	All levels present.
State × Race (5 levels)	8 × 5	28	28	All levels present.
State × Hispanicity	8 × 2	7	7	All levels present.
State × Gender	8 × 2	7	7	All levels present.
Three-Factor Effects		**202**	**160**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	8	Coll. (4,2,1) & (4,3,1); sing. Coll. (5,2,1) & (5,3,1); zero.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	9	Coll. (5,2,1) & (5,3,1); conv.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (1,2,1) & (1,3,1); zero.
State × Age × Race (3 levels)	8 × 6 × 3	70	37	Col (1,5,2) & (1,5,3), (2,4,2) & (2,4,3), (3,2,2) & (3,2,3), (6,4,2) & (6,4,3), (7,4,2) & (7,4,3); sing. Col (2,3,2) & (2,3,3), (2,5,2) & (2,5,3), (3,3,2) & (3,3,3), (3,4,2) & (3,4,3), (3,5,2) & (3,5,3), (5,5,2) & (5,5,3), (6,3,2) & (6,3,3), (6,5,2) & (6,5,3), (7,3,2) & (7,3,3), (7,5,2) & (7,5,3); zero. Drop (2,5,*); conv. Coll. (1,1,1) & (1,1,2), (1,2,1) & (1,2,2), (1,3,1) & (1,3,2), (1,4,1) & (1,4,2), (2,1,1) & (2,1,2), (2,2,1) & (2,2,2), (4,1,2) & (4,1,3), (4,2,2) & (4,2,3), (4,3,2) & (4,3,3), (4,4,2) & (4,4,3), (4,5,2) & (4,5,3), (5,1,2) & (5,1,3), (5,2,2) & (5,2,3), (5,3,2) & (5,3,3), (5,4,2) & (5,4,3), (6,1,2) & (6,1,3), (6,2,2) & (6,2,3); conv.
State × Age × Hispanicity	8 × 6 × 2	35	32	Drop (2,5,1), (3,5,1) & (6,5,1); sing.
State × Age × Gender	8 × 6 × 2	35	35	All levels present.
State × Race (3 levels) × Hispanicity	8 × 3 × 2	14	12	Coll. (2,2,1) & (2,3,1); zero. Coll. (6,2,1) & (6,3,1); sing.
State × Race (3 levels) × Gender	8 × 3 × 2	14	14	All levels present.
State × Hispanicity × Gender	8 × 2 × 2	7	7	All levels present.
Total		**347**	**305**	

Appendix D9: Model Group 9: Pacific

(Alaska, California, Hawaii, Oregon, and Washington)

Table D.9a 2012 NSDUH Person Weight GEM Modeling Summary (Model Group 9: Pacific)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# XVAR[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
res.sdu.nr	1.43	0.56	0.10	1.33194	255	(1.09, 1.40)	(1.09, 1.40)
	1.92	2.25	0.11	1.35979	137	(1.00, 3.00)	(1.00, 3.00)
						(1.05, 1.21)	(1.05, 1.20)
res.sdu.ps	1.92	2.25	0.11	1.35973	197	(0.59, 1.37)	(0.59, 1.37)
	2.31	5.32	1.81	1.48571	193	(0.29, 4.50)	(0.30, 4.50)
						(0.79, 2.07)	(0.79, 2.07)
sel.per.ps	3.70	8.17	2.51	2.71765	287	(0.49, 2.95)	(0.49, 2.95)
	1.97	4.31	1.38	2.76633	262	(0.20, 4.67)	(0.20, 4.67)
						(0.75, 1.32)	(0.75, 1.31)
res.per.nr	2.17	5.09	1.60	2.78828	287	(1.00, 2.90)	(1.00, 2.90)
	1.77	5.40	1.25	3.28420	247	(1.00, 3.60)	(1.00, 3.60)
						(1.40, 1.50)	(1.40, 1.50)
res.per.ps	1.87	5.74	1.52	3.28420	227	(0.20, 1.10)	(0.20, 1.10)
	0.72	2.76	0.57	3.40305	199	(0.20, 4.80)	(0.20, 4.80)
						(0.97, 1.03)	(1.01, 1.01)

[1] For a key to modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[3] Number of proposed covariates (XVAR) on top line and number finalized after modeling.
[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the generalized exponential model (GEM) adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table D.9b Distribution of Weight Adjustment Factors and Weight Products for the 2012 NSDUH Person Weight (Model Group 9: Pacific)

	sel.sdu.des[1]	*res.sdu.nr*[1]		*res.sdu.ps*[1]		*sel.per.des*[1]		*sel.per.ps*[1]		*res.per.nr*[1]		*res.per.ps*[1]	
	1-7[2]	**8**[3]	**1-8**[3]	**9**[4]	**1-9**[4]	**11**[5]	**1-11**[5]	**12**[5]	**1-12**[5]	**13**[6]	**1-13**[6]	**14**[6]	**1-14**[6]
Minimum	101	0.84	103	0.30	39	1.01	66	0.20	28	0.54	33	0.07	13
1%	106	1.02	116	0.57	114	1.01	132	0.39	109	0.98	124	0.20	103
5%	109	1.05	128	0.76	140	1.01	185	0.63	181	1.03	220	0.21	207
10%	127	1.07	142	0.87	157	1.01	278	0.73	276	1.07	337	0.65	313
25%	214	1.10	280	0.99	299	1.18	1,012	0.86	974	1.14	1,160	0.94	946
Median	1,213	1.15	1,296	1.09	1,316	1.45	2,119	0.98	2,122	1.26	2,598	1.02	2,638
75%	1,375	1.24	1,607	1.18	1,801	5.68	6,214	1.11	5,402	1.43	6,258	1.17	5,886
90%	1,448	1.30	1,792	1.35	2,087	10.33	11,874	1.26	11,805	1.64	15,760	1.37	16,471
95%	1,472	1.32	1,858	1.52	2,300	11.43	18,245	1.42	17,233	1.83	22,628	1.39	23,274
99%	1,511	1.60	2,095	2.37	3,138	13.47	24,331	1.73	26,599	2.33	41,919	1.89	40,236
Maximum	2,999	3.00	3,494	4.50	8,444	23.88	75,644	4.67	82,766	3.60	112,786	4.80	133,926
n	17,500	14,830	14,830	14,829	14,829	9,523	9,523	9,523	9,523	7,226	7,226	7,226	7,226
Max/Mean	3.28	-	3.24	-	6.98	-	16.65	-	18.69	-	19.33	-	22.95

Note 1: Weight component 10 and weight products 1-10 are excluded because weight 10 = 1 for all selected dwelling units.
Note 2: Weight component 15 and weight products 1-15 are excluded because weight 15 = 1 for all respondents.
Note 3: Under the generalized exponential model (GEM), nonresponse adjustment factors (weight components #8 and #13) could be less than 1 due to the built-in control for extreme values. For an explanation, see Chapter 2.

[1] Sel.sdu.des refers to selected screener dwelling unit design weight, and sel.per.des refers to selected person design weight. For a key to other modeling abbreviations, see Chapter 5, Exhibit 5.1.
[2] Based on eligible dwelling units.
[3] Based on screener-complete dwelling units.
[4] Based on screener-complete dwelling units, occupants verified eligible.
[5] Based on selected persons.
[6] Based on questionnaire-complete persons.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Model Group 9 Overview

Dwelling Unit Nonresponse

All 23 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the Rent/Housing × percent Black or African American, State × Population Density, State × Group Quarter, State × percent Black or African American, State × percent Hispanic or Latino, State × percent Owner-Occupied, and State × Rent/Housing interactions. Out of 104 proposed variables, 76 were included in the model.

Variable collapsing or dropping was present in all three-factor effects. Out of 128 proposed variables, 38 were included in the model.

In the final model, a total of 137 variables were included; see Exhibit D9.1.

Dwelling Unit Poststratification

All 18 proposed one-factor effects were included in the model.

All 73 proposed two-factor effects were included in the model.

For the three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity and Race × Hispanicity × Gender interactions. Out of 106 proposed variables, 102 were included in the model.

In the final model, a total of 193 variables were included; see Exhibit D9.2.

Selected Person-Level Poststratification

All 36 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the Rent/Housing × percent Black or African American, State × percent Black or African American, State × percent Hispanic or Latino, and State × Rent/Housing interactions. Out of 145 proposed variables, 135 were included in the model.

For three-factor effects, variable collapsing or dropping was present in the Age × Race × Hispanicity, Race × Hispanicity × Gender, State × Age × Race, State × Age × Hispanicity, and State × Race × Hispanicity interactions. Out of 106 proposed variables, 91 were included in the model.

In the final model, a total of 262 variables were included; see Exhibit D9.3.

Respondent Person-Level Nonresponse

All 36 proposed one-factor effects were included in the model.

For two-factor effects, variable collapsing or dropping was present in the Race × Hispanicity, Rent/Housing × percent Black or African American, State × percent Black or African American, State × percent Hispanic or Latino, and State × Rent/Housing interactions. Out of 145 proposed variables, 134 were included in the model.

For three-factor effects, all levels were present for the Age × Race × Gender, Age × Hispanicity × Gender, State × Age × Race, State × Age × Gender, and State × Hispanicity × Gender interactions. All the others were affected by variable collapsing or dropping. Out of 106 proposed variables, 77 were included in the model.

In the final model, a total of 247 variables were included; see Exhibit D9.4.

Respondent Person-Level Poststratification

All 19 proposed one-factor effects were included in the model.

All 81 proposed two-factor effects were included in the model.

For three-factor effects, all levels were present for the Age × Race × Hispanicity, State × Age × Gender, State × Age × Hispanicity, State × Race × Gender, and State × Hispanicity × Gender interactions. All the others were affected by variable collapsing or dropping. Out of 127 proposed variables, 99 were included in the model.

In the final model, a total of 199 variables were included; see Exhibit D9.5.

Exhibit D9.1 Covariates for 2012 NSDUH Person Weights (res.sdu.nr), Model Group 9: Pacific

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**23**	**23**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**104**	**76**	
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (1,1); zero. Coll. (4,1) & (4,2); conv.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Population Density	5 × 4	12	7	Drop (1,1), (2,1), (2,3), (5,3); zero. Drop (2,2); sing.
State × Group Quarter	5 × 3	8	0	Drop (2,2), (3,1), (3,2); zero. Drop (2,1); sing. Drop (1,1), (1,2), (5,1), (5,2); conv.
State × % Black or African American	5 × 3	8	2	Keep (3,2). Coll. (5,1) & (5,2); sing. Drop all others; zero, conv.
State × % Hispanic or Latino	5 × 3	8	4	Drop (1,1), (2,1), (3,1); zero. Drop (1,2); conv.
State × % Owner-Occupied	5 × 3	8	7	Coll. (1,3) & (1,2); conv.
State × Rent/Housing	5 × 5	16	14	Drop (3,3), (3,4); sing.
Three-Factor Effects		**128**	**38**	
State × % Owner-Occupied × % Black or African American	5 × 3 × 3	16	2	Coll. (5,3,1) & (5,3,2), (5,2,1) & (5,2,2); hier. Drop all others; hier./zero.
State × % Owner-Occupied × % Hispanic	5 × 3 × 3	16	6	Keep (2,3,2), (2,2,2), (3,3,2), (3,2,2), (5,3,2), (5,2,2). Drop all others; hier./sing.
State × % Owner-Occupied × Rent/Housing	5 × 3 × 5	32	18	Coll. (1,3,1) & (1,2,1); hier. Coll. (2,3,4) & (2,2,4); sing. Keep (2,3,1), (2,3,2), (2,3,3), (2,2,1), (2,2,2), (2,2,3), (3,2,1), (3,2,2), and (5,*,*), drop all others; hier./zero/sing./conv.
State × Rent/Housing × % Black or African American	5 × 3 × 5	32	3	Keep (1,2,2), (5,2,2). Coll. (5,1,1) & (5,1,2); hier. Drop all others; hier./zero/sing./conv.
State × Rent/Housing × % Hispanic or Latino	5 × 3 × 5	32	9	Keep (1,2,2), (2,1,2), (2,2,2), (3,1,2), (3,2,2), (5,1,2), (5,2,2), (5,3,2), (5,4,2), drop all others; hier./zero/sing.
Total		**255**	**137**	

Exhibit D9.2 Covariates for 2012 NSDUH Person Weights (res.sdu.ps), Model Group 9: Pacific

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**18**	**18**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**73**	**73**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
Three-Factor-Effects		**106**	**102**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (1,2,1) & (1,3,1), (2,2,1) & (2,3,1), (3,2,1) & (3,3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	5 × 5 × 3	32	32	All levels present.
State × Age × Hispanicity	5 × 5 × 2	16	16	All levels present.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	8	All levels present.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**197**	**193**	

Exhibit D9.3 Covariates for 2012 NSDUH Person Weights (sel.per.ps), Model Group 9: Pacific

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**36**	**36**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**145**	**135**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (1,1); zero. Drop (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
State × % Black or African American	5 × 3	8	4	Drop (1,1), (2,1), (3,1); zero. Coll. (5,1) & (5,2); conv.
State × % Hispanic or Latino	5 × 3	8	5	Drop (1,1), (2,1), (3,1); zero.
State × % Owner-Occupied	5 × 3	8	8	All levels present.
State × Rent/Housing	5 × 5	16	15	Drop (3,4); sing.
Three-Factor Effects		**106**	**91**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	5	Coll. (3,2/3,1) & (4,2/3,1); conv.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	1	Coll. (2,1,1) & (3,1,1); conv.
State × Age × Race (3 levels)	5 × 5 × 3	32	31	Coll. (2,4,2) & (2,4,3); sing.
State × Age × Hispanicity	5 × 5 × 2	16	14	Drop (1,4,1), (2,4,1); conv.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	0	Drop all levels; conv.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**287**	**262**	

Exhibit D9.4 Covariates for 2012 NSDUH Person Weights (res.per.nr), Model Group 9: Pacific

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**36**	**36**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	5	4	4	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Relation to Householder	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Rent/Housing	5	4	4	All levels present.
Two-Factor Effects		**145**	**134**	
Age × Race (3 levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	1	Coll. (2,1) & (3,1); conv.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Rent/Housing × % Black or African American	3 × 5	8	6	Drop (1,1); zero. Drop (4,1); sing.
Rent/Housing × % Hispanic or Latino	3 × 5	8	8	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 5	16	16	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
State × % Black or African American	5 × 3	8	4	Drop (1,1), (2,1), (3,1); zero. Drop (5,1); sing.
State × % Hispanic or Latino	5 × 3	8	5	Drop (1,1), (2,1), (3,1); zero.
State × % Owner-Occupied	5 × 3	8	8	All levels present.
State × Rent/Housing	5 × 5	16	15	Drop (3,4); sing.
Three-Factor Effects		**106**	**77**	
Age × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (1,2,1) & (1,3,1), repeat for all age levels; hier.
Age × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
Age × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	0	Coll. (1,2,1) & (1,3,1); hier. Drop (1,2/3,1); conv.
State × Age × Race (3 levels)	5 × 5 × 3	32	16	Coll. (1,1,2) & (1,1,3), repeat for all States and all age levels; conv.
State × Age × Hispanicity	5 × 5 × 2	16	16	All levels present.
State × Age × Gender	5 × 5 × 2	16	16	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (1,2,1) & (1,3,1), repeat for all States; hier.
State × Race (3 levels) × Gender	5 × 3 × 2	8	5	Coll. (2,2,1) & (2,3,1), (3,2,1) & (3,3,1), (5,2,1) & (5,3,1); conv.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**287**	**247**	

Exhibit D9.5 Covariates for 2012 NSDUH Person Weights (res.per.ps), Model Group 9: Pacific

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**19**	**19**	
Intercept	1	1	1	All levels present.
State	5	4	4	All levels present.
Quarter	4	3	3	All levels present.
Age	6	5	5	All levels present.
Race (5 levels)	5	4	4	All levels present.
Gender	2	1	1	All levels present.
Hispanicity	2	1	1	All levels present.
Two-Factor Effects		**81**	**81**	
Age × Race (3 levels)	6 × 3	10	10	All levels present.
Age × Hispanicity	6 × 2	5	5	All levels present.
Age × Gender	6 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Quarter	5 × 4	12	12	All levels present.
State × Age	5 × 6	20	20	All levels present.
State × Race (5 levels)	5 × 5	16	16	All levels present.
State × Hispanicity	5 × 2	4	4	All levels present.
State × Gender	5 × 2	4	4	All levels present.
Three-Factor Effects		**127**	**99**	
Age × Race (3 levels) × Hispanicity	6 × 3 × 2	10	8	Coll. (5,2,1) & (5,3,1); sing. Coll. (3,2,1) & (3,3,1); conv.
Age × Race (3 levels) × Gender	6 × 3 × 2	10	10	All levels present.
Age × Hispanicity × Gender	6 × 2 × 2	5	5	All levels present.
Race (3 levels) × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State × Age × Race (3 levels)	5 × 6 × 3	40	20	Coll. (1,1,2) & (1,1,3), repeat for all age levels and all States; zero/sing./conv.
State × Age × Hispanicity	5 × 6 × 2	20	18	Drop (2,5,1), (3,5,1); conv.
State × Age × Gender	5 × 6 × 2	20	20	All levels present.
State × Race (3 levels) × Hispanicity	5 × 3 × 2	8	4	Coll. (1,2,1) & (1,3,1), repeat for all States; conv.
State × Race (3 levels) × Gender	5 × 3 × 2	8	8	All levels present.
State × Hispanicity × Gender	5 × 2 × 2	4	4	All levels present.
Total		**227**	**199**	

Appendix E: Evaluation of Calibration Weights: Response Rates

Table E.1 2012 NSDUH Weighted Response Rates: United States, District of Columbia, and the 50 States

	Dwelling Unit (DU)					Person Level		Interview Response Rate	
Domain	**Selected DUs**	**Eligible DUs**	**Completed DUs**	**Eligibility Rate**	**Screening Rate**	**Selected Persons**	**Respondents**	**Weight 1-11[1]**	**Weight 1-12[2]**
United States	214,274	178,586	153,873	83.43%	86.07%	87,656	68,309	73.04%	72.92%
Alabama	3,012	2,372	2,141	78.65%	90.30%	1,145	901	74.57%	74.85%
Alaska	2,424	1,869	1,642	76.83%	87.82%	1,076	829	73.34%	74.46%
Arizona	2,771	2,143	1,928	76.44%	90.16%	1,139	922	77.11%	77.23%
Arkansas	2,776	2,292	2,090	82.70%	90.92%	1,212	913	69.77%	69.87%
California	9,489	8,314	6,852	86.08%	82.37%	4,779	3,608	70.20%	70.18%
Colorado	3,071	2,579	2,201	83.17%	85.23%	1,188	927	74.95%	75.03%
Connecticut	2,855	2,535	2,107	88.73%	82.76%	1,261	964	72.36%	72.51%
Delaware	2,847	2,292	2,008	78.08%	87.57%	1,110	893	79.90%	79.33%
District of Columbia	5,055	4,104	3,327	80.88%	80.90%	1,125	962	80.64%	81.16%
Florida	12,768	10,055	8,516	75.81%	84.67%	4,579	3,544	70.57%	70.28%
Georgia	2,365	2,042	1,796	86.31%	87.94%	1,144	885	73.07%	72.71%
Hawaii	3,212	2,761	2,239	85.80%	80.80%	1,285	938	68.98%	69.33%
Idaho	2,300	1,939	1,821	84.78%	93.92%	1,136	921	78.38%	77.54%
Illinois	11,385	9,964	7,678	87.57%	77.04%	4,871	3,672	70.95%	70.80%
Indiana	2,491	2,110	1,921	84.55%	91.01%	1,171	911	72.95%	72.34%
Iowa	2,529	2,199	2,022	86.56%	91.72%	1,137	900	74.74%	75.39%
Kansas	2,598	2,198	1,977	84.94%	89.98%	1,109	912	77.88%	77.14%
Kentucky	2,852	2,407	2,202	84.44%	91.46%	1,184	927	73.49%	73.71%
Louisiana	2,741	2,143	1,977	77.93%	92.28%	1,100	901	77.61%	77.43%
Maine	3,866	2,858	2,585	73.00%	90.56%	1,134	938	79.20%	79.85%
Maryland	2,680	2,308	1,802	86.18%	78.13%	1,074	874	75.90%	75.80%
Massachusetts	3,064	2,653	2,208	85.67%	83.22%	1,253	955	71.52%	72.93%
Michigan	11,441	9,207	7,826	79.39%	85.05%	4,606	3,655	75.75%	75.88%
Minnesota	2,483	2,160	1,975	85.99%	91.57%	1,092	902	81.16%	80.80%
Mississippi	2,553	2,087	1,951	81.96%	93.50%	1,100	901	78.58%	77.90%

(continued)

Table E.1 2012 NSDUH Weighted Response Rates: United States, District of Columbia, and the 50 States (continued)

Domain	Dwelling Unit (DU)					Person Level		Interview Response Rate	
	Selected DUs	Eligible DUs	Completed DUs	Eligibility Rate	Screening Rate	Selected Persons	Respondents	Weight 1-11[1]	Weight 1-12[2]
Missouri	2,879	2,409	2,188	83.62%	90.88%	1,149	915	74.36%	73.05%
Montana	3,295	2,610	2,415	78.09%	92.62%	1,109	876	77.46%	78.41%
Nebraska	2,556	2,175	2,018	85.07%	92.74%	1,170	940	73.14%	73.87%
Nevada	2,354	1,879	1,721	79.87%	91.75%	1,134	903	75.62%	75.72%
New Hampshire	2,990	2,507	2,191	83.88%	87.40%	1,259	950	73.08%	73.72%
New Jersey	2,622	2,227	1,935	84.91%	86.87%	1,155	898	73.64%	73.08%
New Mexico	2,771	2,052	1,889	73.39%	92.22%	1,101	879	74.17%	74.34%
New York	14,547	12,547	9,115	85.42%	71.89%	5,267	3,680	64.38%	63.25%
North Carolina	2,848	2,246	1,990	76.21%	88.48%	1,117	917	75.46%	75.47%
North Dakota	3,374	2,633	2,461	77.65%	93.42%	1,156	895	73.47%	72.75%
Ohio	11,722	10,122	9,023	86.35%	89.14%	4,827	3,687	72.73%	72.72%
Oklahoma	2,960	2,382	2,173	79.51%	91.22%	1,189	908	72.38%	71.15%
Oregon	2,547	2,250	2,019	88.49%	89.57%	1,165	923	76.48%	77.60%
Pennsylvania	11,907	10,256	8,453	85.02%	82.09%	4,705	3,580	70.67%	70.47%
Rhode Island	2,620	2,190	1,957	83.68%	89.37%	1,131	923	77.76%	76.59%
South Carolina	3,306	2,666	2,374	80.44%	88.97%	1,171	938	75.13%	74.56%
South Dakota	2,636	2,163	2,031	81.98%	93.92%	1,113	878	76.12%	76.24%
Tennessee	2,532	2,095	1,929	83.01%	91.91%	1,105	927	81.06%	80.99%
Texas	9,048	7,651	6,792	84.75%	88.52%	4,612	3,625	73.36%	73.52%
Utah	1,793	1,558	1,474	86.99%	94.67%	1,099	926	83.26%	83.60%
Vermont	3,292	2,637	2,317	78.85%	87.81%	1,136	885	73.81%	73.76%
Virginia	2,576	2,293	2,027	88.97%	88.47%	1,095	894	76.50%	76.10%
Washington	2,700	2,306	2,078	85.67%	90.10%	1,218	928	71.82%	71.37%
West Virginia	3,222	2,675	2,399	82.94%	89.39%	1,217	976	74.07%	74.74%
Wisconsin	2,440	2,041	1,890	83.27%	92.37%	1,098	875	75.55%	76.01%
Wyoming	3,109	2,425	2,222	77.59%	91.72%	1,148	928	77.48%	77.61%

[1] Includes DU-level and person-level design weights, DU nonresponse adjustment, and DU poststratification adjustment.

[2] Includes a selected person poststratification weight.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Appendix F: Evaluation of Calibration Weights: Dwelling Unit–Level Percentages of Extreme Weights and Outwinsors

Table F.1 2012 NSDUH Dwelling Unit–Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States

		Before nr[1] (WT1*...*WT7)			After nr[1] & Before ps[2] (WT1*...*WT8)			After ps[2] (WT1*...*WT9)		
Domain	***n***	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**
United States	153,873	3.51%	3.38%	0.44%	2.74%	3.32%	0.42%	1.80%	3.63%	1.02%
Alabama	2,141	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.98%	2.09%	0.56%
Alaska	1,642	0.37%	0.71%	0.19%	1.77%	2.54%	0.21%	2.07%	4.10%	0.76%
Arizona	1,928	10.27%	12.71%	1.78%	4.88%	5.69%	0.41%	1.04%	2.50%	0.52%
Arkansas	2,090	2.97%	2.32%	0.18%	0.00%	0.00%	0.00%	0.77%	1.69%	0.41%
California	6,852	0.20%	0.36%	0.07%	2.45%	2.81%	0.28%	2.32%	5.49%	2.08%
Colorado	2,201	9.86%	20.08%	8.00%	7.95%	16.36%	3.07%	1.54%	3.36%	0.64%
Connecticut	2,107	0.00%	0.00%	0.00%	2.85%	6.20%	0.84%	1.33%	4.06%	1.40%
Delaware	2,008	1.29%	2.14%	0.54%	1.20%	1.45%	0.13%	1.34%	3.77%	1.25%
District of Columbia	3,327	11.78%	10.24%	1.12%	2.34%	4.16%	1.40%	1.53%	3.17%	0.44%
Florida	8,516	4.78%	6.35%	0.98%	4.78%	5.39%	0.47%	1.30%	2.37%	0.50%
Georgia	1,796	0.39%	0.73%	0.09%	0.56%	1.01%	0.26%	1.50%	3.29%	0.81%
Hawaii	2,239	8.58%	12.01%	2.06%	2.90%	2.86%	0.19%	1.92%	4.60%	0.89%
Idaho	1,821	0.22%	0.02%	0.04%	0.00%	0.00%	0.00%	1.81%	3.72%	1.09%
Illinois	7,678	0.46%	0.39%	0.05%	4.69%	6.73%	1.14%	0.85%	1.60%	0.44%
Indiana	1,921	0.21%	0.05%	0.00%	0.21%	0.74%	0.46%	2.50%	4.19%	0.98%
Iowa	2,022	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	1.04%	2.34%	0.49%
Kansas	1,977	16.49%	19.45%	1.76%	1.21%	1.73%	0.27%	3.74%	6.95%	2.13%
Kentucky	2,202	6.63%	8.69%	0.19%	4.45%	5.86%	0.40%	0.77%	1.30%	0.18%
Louisiana	1,977	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.76%	1.70%	0.34%
Maine	2,585	12.03%	13.71%	0.93%	9.17%	10.51%	0.92%	4.64%	3.74%	1.22%
Maryland	1,802	7.82%	7.90%	0.20%	3.66%	3.98%	0.35%	2.39%	3.97%	0.81%
Massachusetts	2,208	4.03%	4.39%	0.30%	4.08%	5.57%	1.03%	5.12%	7.03%	2.30%
Michigan	7,826	0.88%	1.13%	0.12%	0.00%	0.00%	0.00%	0.41%	0.86%	0.20%
Minnesota	1,975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.81%	1.89%	0.38%
Mississippi	1,951	0.26%	0.28%	0.01%	0.05%	0.02%	0.00%	1.74%	1.59%	0.38%

(continued)

Table F.1 2012 NSDUH Dwelling Unit–Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States (continued)

		Before nr[1] (WT1*...*WT7)			After nr[1] & Before ps[2] (WT1*...*WT8)			After ps[2] (WT1*...*WT9)		
Domain	***n***	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**
Missouri	2,188	0.00%	0.00%	0.00%	2.19%	2.43%	0.23%	1.74%	3.55%	0.70%
Montana	2,415	1.41%	2.69%	1.05%	2.61%	3.49%	0.30%	2.32%	3.70%	0.71%
Nebraska	2,018	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.76%	8.48%	2.01%
Nevada	1,721	3.20%	3.11%	0.12%	2.79%	2.63%	0.31%	2.61%	6.68%	1.60%
New Hampshire	2,191	0.50%	0.98%	0.33%	1.19%	1.44%	0.12%	5.34%	6.03%	1.57%
New Jersey	1,935	0.00%	0.00%	0.00%	0.36%	1.24%	0.44%	1.19%	3.17%	0.80%
New Mexico	1,889	10.27%	12.12%	1.19%	8.79%	10.48%	0.90%	2.54%	4.56%	1.30%
New York	9,115	5.83%	6.15%	0.79%	1.97%	3.62%	0.80%	2.02%	5.99%	2.42%
North Carolina	1,990	4.17%	4.42%	0.58%	1.31%	2.11%	0.09%	2.56%	5.89%	1.40%
North Dakota	2,461	4.35%	5.13%	0.17%	1.67%	2.08%	0.07%	2.76%	5.71%	1.36%
Ohio	9,023	11.93%	13.98%	0.74%	9.09%	10.57%	0.64%	1.01%	1.39%	0.26%
Oklahoma	2,173	1.84%	2.05%	0.23%	0.64%	1.43%	0.52%	1.24%	2.77%	0.39%
Oregon	2,019	0.00%	0.00%	0.00%	1.14%	1.62%	0.03%	3.22%	5.49%	1.47%
Pennsylvania	8,453	0.44%	0.76%	0.04%	1.41%	3.02%	0.78%	0.62%	1.67%	0.45%
Rhode Island	1,957	0.00%	0.00%	0.00%	1.43%	1.73%	0.23%	1.48%	4.95%	2.08%
South Carolina	2,374	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.53%	3.65%	0.54%
South Dakota	2,031	0.00%	0.00%	0.00%	0.84%	1.59%	0.15%	1.77%	3.66%	0.90%
Tennessee	1,929	3.16%	2.80%	0.53%	0.00%	0.00%	0.00%	4.20%	6.63%	0.91%
Texas	6,792	1.84%	2.14%	0.14%	2.77%	3.18%	0.26%	1.43%	2.52%	0.90%
Utah	1,474	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.58%	3.17%	0.44%
Vermont	2,317	14.98%	21.18%	2.49%	11.48%	16.42%	2.06%	1.17%	2.64%	0.60%
Virginia	2,027	0.25%	0.10%	0.07%	2.66%	3.05%	0.30%	2.96%	4.61%	0.84%
Washington	2,078	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.07%	4.62%	1.18%
West Virginia	2,399	0.17%	0.05%	0.01%	0.00%	0.00%	0.00%	0.50%	1.32%	0.20%
Wisconsin	1,890	1.96%	1.74%	0.29%	3.23%	3.58%	0.21%	2.43%	3.44%	0.57%
Wyoming	2,222	0.00%	0.00%	0.00%	1.40%	1.41%	0.18%	3.24%	3.53%	0.93%

[1] nr = nonresponse adjustment.

[2] ps = poststratification adjustment.

[3] Weighted extreme value percentage = $100*\sum_k w_{ek} / \sum_k w_k$, where w_{ek} denotes the weight for extreme weights and w_k denotes the weight for both extreme weights and nonextreme weights.

[4] Outwinsor weight percentage = $100*\sum_k (w_{ek} - b_k) / \sum_k w_k$, where b_k denotes the cutoff point for defining the extreme weight.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Appendix G: Evaluation of Calibration Weights: Person-Level Percentages of Extreme Weights and Outwinsors

Table G.1 2012 NSDUH Selected Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States

		Before sel.per.ps[1] (WT1*...*WT11)			After sel.per.ps[1] (WT1*...*WT12)		
Domain	***n***	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**
United States	87,656	3.41%	7.03%	1.86%	1.97%	4.37%	1.06%
Alabama	1,145	2.01%	4.27%	0.92%	1.57%	2.04%	0.38%
Alaska	1,076	5.30%	11.49%	2.72%	2.60%	5.21%	1.75%
Arizona	1,139	3.86%	11.35%	2.57%	2.81%	4.31%	0.53%
Arkansas	1,212	3.55%	4.31%	1.35%	2.39%	7.10%	1.98%
California	4,779	3.41%	8.20%	2.62%	1.76%	3.99%	1.31%
Colorado	1,188	1.94%	3.89%	0.72%	1.77%	4.71%	1.24%
Connecticut	1,261	2.54%	7.54%	2.58%	2.62%	8.13%	2.00%
Delaware	1,110	3.42%	6.74%	1.71%	1.26%	3.42%	1.19%
District of Co	1,125	3.02%	3.46%	0.44%	1.16%	2.27%	0.51%
Florida	4,579	3.04%	6.25%	1.13%	1.16%	1.79%	0.61%
Georgia	1,144	3.76%	7.27%	1.58%	2.10%	4.00%	0.95%
Hawaii	1,285	3.89%	8.70%	1.94%	2.02%	5.20%	1.63%
Idaho	1,136	4.31%	7.91%	2.16%	2.46%	4.98%	1.61%
Illinois	4,871	2.83%	4.85%	1.01%	2.03%	4.49%	0.81%
Indiana	1,171	2.99%	7.27%	2.03%	1.79%	4.11%	0.74%
Iowa	1,137	2.73%	4.91%	1.76%	2.55%	5.20%	1.44%
Kansas	1,109	4.33%	7.92%	2.30%	1.98%	4.08%	0.87%
Kentucky	1,184	3.80%	6.14%	1.25%	1.86%	1.80%	0.39%
Louisiana	1,100	2.09%	2.79%	0.60%	1.91%	5.11%	0.69%
Maine	1,134	6.17%	6.70%	2.18%	2.38%	3.09%	0.65%
Maryland	1,074	2.14%	3.60%	0.63%	2.05%	4.06%	1.02%
Massachusetts	1,253	3.35%	7.21%	2.01%	2.63%	7.76%	2.21%
Michigan	4,606	2.08%	3.04%	0.66%	1.22%	2.32%	0.47%
Minnesota	1,092	2.29%	5.15%	1.09%	2.20%	5.39%	1.44%
Mississippi	1,100	4.09%	4.56%	0.78%	0.82%	1.95%	0.66%

(continued)

Table G.1 2012 NSDUH Selected Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States (continued)

		Before sel.per.ps[1] (WT1*...*WT11)			After sel.per.ps[1] (WT1*...*WT12)		
Domain	***n***	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**	**% Unweighted**	**% Weighted[2]**	**% Outwinsor[3]**
Missouri	1,149	2.79%	5.43%	1.29%	3.31%	8.78%	2.44%
Montana	1,109	3.34%	4.74%	0.97%	1.89%	3.30%	0.99%
Nebraska	1,170	2.31%	5.14%	1.59%	2.56%	8.16%	2.90%
Nevada	1,134	4.23%	8.60%	2.39%	2.65%	4.71%	1.09%
New Hampshire	1,259	3.26%	4.55%	1.09%	1.27%	2.10%	0.61%
New Jersey	1,155	1.82%	2.91%	0.56%	1.73%	3.21%	0.67%
New Mexico	1,101	3.54%	7.39%	1.53%	1.18%	2.53%	0.66%
New York	5,267	3.78%	10.31%	3.74%	3.15%	9.03%	2.52%
North Carolina	1,117	2.69%	4.95%	1.03%	1.79%	5.22%	0.98%
North Dakota	1,156	3.72%	7.54%	1.49%	1.38%	2.53%	0.75%
Ohio	4,827	4.10%	5.65%	0.97%	1.70%	2.54%	0.31%
Oklahoma	1,189	2.35%	4.29%	0.75%	1.85%	2.31%	0.33%
Oregon	1,165	4.98%	10.15%	3.09%	2.23%	5.12%	1.83%
Pennsylvania	4,705	3.02%	6.97%	1.52%	1.68%	4.05%	0.52%
Rhode Island	1,131	2.48%	6.72%	3.28%	2.21%	5.57%	1.62%
South Carolina	1,171	3.42%	6.09%	0.98%	0.94%	1.88%	0.63%
South Dakota	1,113	3.77%	9.30%	2.06%	2.25%	3.85%	0.63%
Tennessee	1,105	4.71%	9.81%	2.79%	1.36%	2.57%	0.44%
Texas	4,612	5.18%	11.19%	3.36%	2.91%	4.71%	1.02%
Utah	1,099	4.28%	7.66%	1.48%	0.82%	1.73%	0.32%
Vermont	1,136	6.25%	11.64%	2.89%	3.17%	7.27%	1.12%
Virginia	1,095	3.20%	7.44%	1.27%	1.46%	3.73%	0.53%
Washington	1,218	3.69%	7.74%	2.25%	2.71%	6.25%	1.59%
West Virginia	1,217	1.15%	1.46%	0.34%	0.90%	1.55%	0.57%
Wisconsin	1,098	3.10%	5.99%	1.66%	2.28%	3.71%	0.62%
Wyoming	1,148	3.40%	4.75%	1.62%	1.92%	4.40%	1.30%

[1] Before sel.per.ps (WT1*...*WT11) and after sel.per.ps (WT1*...*WT12) used demographic variables from screener data for all selected persons; ps = poststratification adjustment.

[2] Weighted extreme value percentage = $100*\sum_k w_{ek} / \sum_k w_k$, where w_{ek} denotes the weight for extreme weights and w_k denotes the weight for both extreme weights and nonextreme weights.

[3] Outwinsor weight percentage = $100*\sum_k (w_{ek} - b_k) / \sum_k w_k$, where b_k denotes the cutoff point for defining the extreme weight.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table G.2 2012 NSDUH Respondent Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States

Domain	*n*	Before res.per.nr[1] (WT1*…*WT12)			After res.per.nr[1] (WT1*…*WT13)			Before res.per.ps[2] (WT1*…*WT13)			After res.per.ps[2] (WT1*…*WT14)		
		% Unweighted	% Weighted[3]	% Outwinsor[4]	% Unweighted	% Weighted[3]	% Outwinsor[4]	% Unweighted	% Weighted[3]	% Outwinsor[4]	% Unweighted	% Weighted[3]	% Outwinsor[4]
United States	68,309	2.02%	4.54%	1.15%	1.71%	4.90%	1.06%	1.76%	4.98%	1.15%	1.04%	3.24%	0.66%
Alabama	901	1.89%	2.61%	0.46%	1.44%	2.17%	0.37%	1.44%	2.17%	0.34%	1.66%	2.65%	0.44%
Alaska	829	2.41%	5.49%	2.16%	2.77%	5.65%	1.21%	2.65%	5.78%	1.41%	1.81%	3.51%	0.28%
Arizona	922	2.60%	4.09%	0.49%	1.95%	2.42%	0.28%	1.74%	2.35%	0.32%	1.08%	1.74%	0.23%
Arkansas	913	2.41%	4.91%	1.17%	3.50%	7.88%	1.49%	3.50%	7.88%	1.48%	1.53%	5.46%	0.90%
California	3,608	1.94%	4.86%	1.56%	1.61%	5.91%	1.33%	1.77%	6.29%	1.67%	0.53%	3.24%	0.72%
Colorado	927	2.05%	4.75%	1.44%	1.83%	5.02%	0.91%	1.62%	4.14%	1.10%	1.19%	2.27%	0.31%
Connecticut	964	2.59%	8.80%	2.07%	1.66%	7.72%	2.16%	1.56%	7.59%	2.12%	1.35%	4.06%	0.54%
Delaware	893	1.23%	3.40%	1.01%	1.12%	3.79%	0.62%	1.23%	3.87%	0.66%	0.78%	4.13%	0.53%
District of Co	962	1.14%	2.58%	0.57%	0.73%	2.05%	0.22%	0.83%	2.19%	0.25%	1.35%	3.86%	0.55%
Florida	3,544	1.24%	2.05%	0.70%	1.35%	3.74%	0.64%	1.50%	4.00%	0.68%	0.23%	0.31%	0.02%
Georgia	885	2.03%	3.30%	0.94%	3.16%	6.50%	2.27%	2.94%	6.30%	2.17%	2.49%	6.91%	1.68%
Hawaii	938	2.13%	5.67%	1.81%	1.71%	6.65%	1.06%	1.71%	6.65%	1.26%	0.32%	0.37%	0.03%
Idaho	921	2.93%	6.09%	1.87%	2.61%	6.20%	1.90%	2.61%	5.55%	1.83%	2.50%	5.65%	1.15%
Illinois	3,672	1.91%	4.19%	0.75%	0.98%	3.66%	0.68%	1.12%	3.90%	0.76%	0.25%	1.07%	0.33%
Indiana	911	1.76%	3.87%	0.86%	1.21%	2.57%	0.15%	1.10%	2.46%	0.15%	1.32%	2.75%	0.99%
Iowa	900	2.67%	5.10%	1.68%	2.67%	6.09%	1.05%	2.56%	5.73%	1.08%	1.67%	6.44%	2.10%
Kansas	912	1.97%	4.46%	1.03%	1.75%	6.97%	1.45%	1.64%	6.52%	1.41%	2.19%	8.23%	2.66%
Kentucky	927	1.73%	1.73%	0.33%	1.40%	1.60%	0.21%	1.73%	2.74%	0.46%	1.40%	3.72%	0.38%
Louisiana	901	1.44%	3.13%	0.61%	1.89%	4.23%	0.76%	1.89%	4.23%	0.81%	0.22%	0.16%	0.01%
Maine	938	2.03%	2.41%	0.53%	1.28%	1.52%	0.36%	1.17%	1.32%	0.25%	1.17%	4.22%	1.22%
Maryland	874	1.83%	4.02%	1.28%	1.83%	5.23%	0.93%	1.72%	5.10%	1.05%	2.06%	4.88%	1.24%
Massachusetts	955	3.98%	11.81%	2.76%	2.62%	7.96%	1.98%	2.41%	7.15%	1.94%	0.63%	2.08%	0.35%
Michigan	3,655	1.12%	1.95%	0.51%	1.09%	3.33%	0.50%	1.09%	3.05%	0.52%	0.30%	1.19%	0.20%
Minnesota	902	1.66%	4.13%	1.02%	1.77%	6.50%	1.74%	1.66%	6.27%	1.58%	1.66%	6.99%	1.48%
Mississippi	901	1.00%	1.99%	0.75%	1.66%	4.05%	1.39%	1.78%	4.51%	1.43%	0.78%	3.65%	0.37%

(continued)

Table G.2 2012 NSDUH Respondent Person-Level Percentages of Extreme Weights and Outwinsors: United States, District of Columbia, and the 50 States (continued)

		Before res.per.nr[1] (WT1*…*WT12)			After res.per.nr[1] (WT1*…*WT13)			Before res.per.ps[2] (WT1*…*WT13)			After res.per.ps[2] (WT1*…*WT14)		
Domain	***n***	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**	**% Unweighted**	**% Weighted[3]**	**% Outwinsor[4]**
Missouri	915	3.28%	10.36%	2.93%	3.17%	9.07%	2.27%	3.61%	9.44%	2.32%	1.53%	5.20%	1.51%
Montana	876	1.71%	3.48%	1.23%	1.48%	3.68%	0.66%	1.37%	3.32%	0.61%	0.91%	1.37%	0.55%
Nebraska	940	2.77%	9.31%	3.55%	2.23%	7.52%	1.90%	2.34%	7.77%	2.13%	1.91%	4.74%	1.47%
Nevada	903	2.88%	5.71%	1.40%	2.88%	6.79%	1.19%	2.88%	6.79%	1.17%	1.00%	2.79%	0.37%
New Hampshire	950	1.58%	2.31%	0.67%	2.11%	3.64%	0.80%	2.42%	4.52%	1.15%	2.42%	4.06%	0.85%
New Jersey	898	1.89%	3.42%	0.72%	1.34%	2.79%	0.63%	1.22%	2.67%	0.61%	1.78%	5.19%	0.68%
New Mexico	879	1.14%	3.08%	0.80%	1.02%	3.01%	1.07%	1.25%	3.38%	1.20%	0.80%	3.40%	0.38%
New York	3,680	3.29%	9.06%	2.76%	3.72%	12.57%	2.80%	3.83%	12.87%	3.23%	1.44%	7.09%	1.21%
North Carolina	917	1.74%	3.48%	0.46%	0.76%	3.07%	0.88%	0.76%	3.07%	0.89%	0.98%	3.06%	1.01%
North Dakota	895	1.79%	3.47%	1.04%	1.23%	3.23%	0.63%	1.23%	3.23%	0.62%	0.67%	2.79%	0.57%
Ohio	3,687	1.82%	2.72%	0.36%	0.98%	2.43%	0.20%	0.95%	2.36%	0.19%	0.52%	0.54%	0.05%
Oklahoma	908	1.65%	2.00%	0.40%	1.10%	1.94%	0.32%	1.21%	2.09%	0.33%	0.88%	3.05%	0.74%
Oregon	923	2.06%	5.17%	2.10%	2.28%	6.79%	1.77%	2.38%	6.88%	1.82%	1.63%	4.25%	0.51%
Pennsylvania	3,580	1.59%	3.89%	0.50%	0.89%	2.95%	0.33%	0.95%	3.10%	0.34%	0.67%	2.39%	0.35%
Rhode Island	923	2.49%	6.54%	1.96%	1.63%	5.87%	1.52%	1.63%	5.87%	1.53%	0.76%	1.03%	0.16%
South Carolina	938	0.96%	2.29%	0.84%	0.75%	1.66%	0.56%	0.75%	1.66%	0.59%	0.11%	0.47%	0.12%
South Dakota	878	2.51%	4.18%	0.54%	1.71%	4.07%	0.94%	1.82%	4.50%	0.96%	1.48%	3.74%	0.80%
Tennessee	927	0.86%	2.11%	0.33%	1.29%	2.88%	0.67%	1.40%	3.13%	0.73%	1.51%	2.69%	0.38%
Texas	3,625	3.09%	5.29%	1.20%	2.04%	3.62%	0.67%	2.21%	3.92%	0.73%	1.16%	1.43%	0.22%
Utah	926	0.86%	1.61%	0.33%	0.65%	3.50%	0.87%	0.65%	3.49%	0.87%	0.86%	2.44%	0.33%
Vermont	885	3.39%	7.40%	1.35%	4.07%	7.25%	1.66%	3.95%	7.36%	1.80%	1.69%	2.60%	0.86%
Virginia	894	2.13%	5.58%	0.98%	1.57%	6.50%	1.42%	1.57%	6.01%	1.35%	2.57%	10.17%	2.44%
Washington	928	2.91%	6.69%	1.57%	1.72%	2.81%	0.73%	1.83%	3.25%	0.80%	1.40%	1.52%	0.15%
West Virginia	976	1.02%	1.48%	0.61%	1.33%	2.79%	0.63%	1.43%	2.44%	0.85%	1.13%	1.69%	0.32%
Wisconsin	875	2.29%	3.96%	0.67%	1.37%	3.08%	0.82%	1.37%	3.08%	0.77%	1.60%	4.20%	0.92%
Wyoming	928	1.94%	4.63%	1.50%	1.40%	3.00%	0.59%	1.62%	3.67%	0.80%	0.65%	1.59%	0.28%

[1] Before res.per.nr (WT1*...*WT12) and after res.per.nr (WT1*...*WT13) used demographic variables from screener data for all respondents; nr = nonresponse adjustment.

[2] Before res.per.ps (WT1*...*WT13) and after res.per.ps (WT1*...*WT14) used demographic variables from questionnaire data for all respondents; ps = poststratification adjustment.

[3] Weighted outlier percentage = $100*\sum_k w_{ok}/\sum_k w_k$, where w_{ok} denotes the weight for outliers and w_k denotes the weight for both outliers and nonoutliers.

[4] Outwinsor weight percentage = $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the cutoff point for defining the extreme weight.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Appendix H: Evaluation of Calibration Weights: Slippage Rates

Table H.1 2012 NSDUH Slippage Rates: UNITED STATES

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		68,309	260,057,325	260,057,325	260,057,325	0.00	0.00
Quarter	**Quarter 1**	16,241	64,783,210	64,783,210	64,783,210	0.00	0.00
	Quarter 2	18,090	64,934,336	64,934,336	64,934,336	0.00	0.00
	Quarter 3	17,836	65,092,650	65,092,650	65,092,650	0.00	0.00
	Quarter 4	16,142	65,247,130	65,247,130	65,247,130	0.00	0.00
Age Group	**12-17**	22,473	24,978,268	24,933,051	24,933,051	0.18	0.00
	18-25	22,529	34,362,630	34,589,953	34,589,953	-0.66	0.00
	26-34	6,484	36,900,291	36,911,543	36,911,543	-0.03	0.00
	35-49	9,076	61,346,701	61,035,041	61,035,041	0.51	0.00
	50-64	4,762	62,584,251	60,724,221	60,724,221	3.06	0.00
	65+	2,985	39,885,185	41,863,516	41,863,516	-4.73	0.00
Race	**White**	50,568	195,721,173	205,426,089	205,426,089	-4.72	0.00
	Black or African American	9,021	33,217,417	32,495,008	32,495,008	2.22	0.00
	Other	8,720	31,118,735	22,136,228	22,136,228	40.58	0.00
Hispanicity	**Hispanic or Latino**	11,483	40,865,453	40,299,676	40,299,676	1.40	0.00
	Non-Hispanic or Latino	56,826	219,191,873	219,757,649	219,757,649	-0.26	0.00
Gender	**Male**	32,861	125,856,460	125,892,900	125,892,900	-0.03	0.00
	Female	35,448	134,200,866	134,164,425	134,164,425	0.03	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.2 2012 NSDUH Slippage Rates: ALABAMA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		901	4,005,432	4,005,432	4,005,432	0.00	0.00
Quarter	**Quarter 1**	206	999,184	999,184	999,184	0.00	0.00
	Quarter 2	224	1,000,517	1,000,517	1,000,517	0.00	0.00
	Quarter 3	272	1,002,072	1,002,072	1,002,072	0.00	0.00
	Quarter 4	199	1,003,659	1,003,659	1,003,659	0.00	0.00
Age Group	**12-17**	279	386,242	384,244	384,244	0.52	0.00
	18-25	308	529,857	536,932	536,932	-1.32	0.00
	26-34	84	524,551	538,091	538,091	-2.52	0.00
	35-49	125	925,187	912,523	912,523	1.39	0.00
	50-64	61	909,177	954,182	954,182	-4.72	0.00
	65+	44	730,419	679,461	679,461	7.50	-0.00
Race	**White**	539	2,809,348	2,853,932	2,853,932	-1.56	0.00
	Black or African American	322	1,006,480	1,024,994	1,024,994	-1.81	0.00
	Other	40	189,604	126,507	126,507	49.88	0.00
Hispanicity	**Hispanic or Latino**	29	124,050	137,969	137,969	-10.09	0.00
	Non-Hispanic or Latino	872	3,881,383	3,867,464	3,867,464	0.36	0.00
Gender	**Male**	425	1,906,473	1,907,564	1,907,564	-0.06	-0.00
	Female	476	2,098,959	2,097,868	2,097,868	0.05	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.3 2012 NSDUH Slippage Rates: ALASKA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		829	577,147	577,147	577,147	-0.00	-0.00
Quarter	**Quarter 1**	239	143,696	143,696	143,696	0.00	0.00
	Quarter 2	210	144,125	144,125	144,125	0.00	0.00
	Quarter 3	230	144,508	144,508	144,508	-0.00	0.00
	Quarter 4	150	144,819	144,819	144,819	0.00	0.00
Age Group	**12-17**	233	60,451	60,308	60,308	0.24	-0.00
	18-25	287	83,454	81,619	81,619	2.25	0.00
	26-34	95	86,667	89,969	89,969	-3.67	0.00
	35-49	130	135,363	136,165	136,165	-0.59	-0.00
	50-64	54	138,649	147,819	147,819	-6.20	0.00
	65+	30	72,563	61,269	61,269	18.43	0.00
Race	**White**	557	402,786	402,098	402,098	0.17	-0.00
	Black or African American	24	18,331	19,007	19,007	-3.56	0.00
	Other	248	156,030	156,042	156,042	-0.01	0.00
Hispanicity	**Hispanic or Latino**	63	33,866	32,109	32,109	5.47	-0.00
	Non-Hispanic or Latino	766	543,281	545,038	545,038	-0.32	0.00
Gender	**Male**	401	288,168	293,291	293,292	-1.75	-0.00
	Female	428	288,978	283,855	283,855	1.80	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.4 2012 NSDUH Slippage Rates: ARIZONA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		922	5,362,657	5,362,657	5,362,657	0.00	0.00
Quarter	**Quarter 1**	276	1,332,896	1,332,896	1,332,896	-0.00	0.00
	Quarter 2	184	1,338,264	1,338,264	1,338,264	0.00	0.00
	Quarter 3	228	1,343,420	1,343,420	1,343,420	0.00	0.00
	Quarter 4	234	1,348,077	1,348,077	1,348,077	-0.00	0.00
Age Group	**12-17**	313	541,550	539,163	539,163	0.44	0.00
	18-25	292	723,118	713,584	713,584	1.34	0.00
	26-34	99	741,139	759,644	759,644	-2.44	0.00
	35-49	124	1,203,994	1,209,340	1,209,340	-0.44	0.00
	50-64	63	1,487,519	1,182,994	1,182,994	25.74	0.00
	65+	31	665,337	957,934	957,934	-30.54	0.00
Race	**White**	707	4,333,865	4,585,263	4,585,263	-5.48	0.00
	Black or African American	54	230,315	226,910	226,910	1.50	0.00
	Other	161	798,477	550,484	550,484	45.05	0.00
Hispanicity	**Hispanic or Latino**	378	1,459,037	1,463,787	1,463,787	-0.32	0.00
	Non-Hispanic or Latino	544	3,903,621	3,898,871	3,898,871	0.12	0.00
Gender	**Male**	444	2,617,812	2,617,137	2,617,137	0.03	0.00
	Female	478	2,744,845	2,745,520	2,745,520	-0.02	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.5 2012 NSDUH Slippage Rates: ARKANSAS

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		913	2,422,926	2,422,926	2,422,926	0.00	0.00
Quarter	**Quarter 1**	199	604,436	604,436	604,436	0.00	0.00
	Quarter 2	282	605,156	605,156	605,156	0.00	-0.00
	Quarter 3	200	606,124	606,124	606,124	0.00	-0.00
	Quarter 4	232	607,210	607,210	607,210	0.00	0.00
Age Group	**12-17**	312	236,048	236,048	236,048	0.00	-0.00
	18-25	304	310,698	317,735	317,735	-2.21	-0.00
	26-34	87	341,525	334,488	334,488	2.10	0.00
	35-49	103	539,208	545,591	545,591	-1.17	0.00
	50-64	58	543,503	562,157	562,157	-3.32	0.00
	65+	49	451,944	426,908	426,908	5.86	0.00
Race	**White**	719	1,912,686	1,972,067	1,972,067	-3.01	0.00
	Black or African American	136	349,460	355,467	355,467	-1.69	0.00
	Other	58	160,780	95,392	95,392	68.55	-0.00
Hispanicity	**Hispanic or Latino**	80	152,363	140,612	140,612	8.36	0.00
	Non-Hispanic or Latino	833	2,270,564	2,282,314	2,282,314	-0.51	-0.00
Gender	**Male**	409	1,169,794	1,169,794	1,169,794	0.00	0.00
	Female	504	1,253,132	1,253,132	1,253,132	0.00	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.6 2012 NSDUH Slippage Rates: CALIFORNIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,608	31,424,054	31,424,054	31,424,054	0.00	0.00
Quarter	**Quarter 1**	946	7,823,059	7,823,059	7,823,059	-0.00	0.00
	Quarter 2	904	7,845,198	7,845,198	7,845,198	0.00	0.00
	Quarter 3	950	7,867,268	7,867,268	7,867,268	-0.00	0.00
	Quarter 4	808	7,888,528	7,888,529	7,888,529	-0.00	0.00
Age Group	**12-17**	1,164	3,166,343	3,139,169	3,139,169	0.87	0.00
	18-25	1,187	4,407,880	4,452,711	4,452,711	-1.01	0.00
	26-34	415	4,892,630	4,802,643	4,802,643	1.87	0.00
	35-49	486	7,761,047	7,649,117	7,649,117	1.46	0.00
	50-64	219	6,830,212	6,873,706	6,873,706	-0.63	0.00
	65+	137	4,365,943	4,506,708	4,506,708	-3.12	0.00
Race	**White**	2,300	20,549,639	23,254,991	23,254,991	-11.63	0.00
	Black or African American	257	2,100,386	2,006,847	2,006,847	4.66	-0.00
	Other	1,051	8,774,029	6,162,216	6,162,216	42.38	0.00
Hispanicity	**Hispanic or Latino**	1,645	11,096,233	11,177,945	11,177,945	-0.73	0.00
	Non-Hispanic or Latino	1,963	20,327,821	20,246,109	20,246,109	0.40	0.00
Gender	**Male**	1,735	15,396,931	15,394,395	15,394,395	0.02	0.00
	Female	1,873	16,027,124	16,029,659	16,029,659	-0.02	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.7 2012 NSDUH Slippage Rates: COLORADO

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		927	4,260,412	4,260,412	4,260,412	0.00	0.00
Quarter	**Quarter 1**	212	1,058,623	1,058,623	1,058,623	0.00	0.00
	Quarter 2	229	1,063,136	1,063,136	1,063,136	0.00	0.00
	Quarter 3	239	1,067,413	1,067,413	1,067,413	0.00	0.00
	Quarter 4	247	1,071,240	1,071,240	1,071,240	0.00	0.00
Age Group	**12-17**	318	397,001	399,087	399,087	-0.52	0.00
	18-25	301	573,180	560,123	560,123	2.33	0.00
	26-34	83	649,195	660,166	660,166	-1.66	0.00
	35-49	111	1,033,849	1,033,849	1,033,849	-0.00	0.00
	50-64	71	1,013,685	1,010,341	1,010,341	0.33	0.00
	65+	43	593,502	596,846	596,846	-0.56	0.00
Race	**White**	738	3,575,613	3,797,172	3,797,172	-5.83	0.00
	Black or African American	43	164,145	170,118	170,118	-3.51	-0.00
	Other	146	520,653	293,122	293,122	77.62	0.00
Hispanicity	**Hispanic or Latino**	250	820,361	803,430	803,430	2.11	0.00
	Non-Hispanic or Latino	677	3,440,050	3,456,982	3,456,982	-0.49	0.00
Gender	**Male**	447	2,103,568	2,103,702	2,103,702	-0.01	0.00
	Female	480	2,156,843	2,156,709	2,156,709	0.01	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.8 2012 NSDUH Slippage Rates: CONNECTICUT

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		964	3,034,241	3,034,241	3,034,241	0.00	0.00
Quarter	**Quarter 1**	251	757,038	757,038	757,038	0.00	0.00
	Quarter 2	201	757,909	757,909	757,909	0.00	0.00
	Quarter 3	248	759,030	759,030	759,030	0.00	0.00
	Quarter 4	264	760,264	760,264	760,264	0.00	0.00
Age Group	**12-17**	289	290,827	289,862	289,862	0.33	0.00
	18-25	330	362,809	373,279	373,279	-2.80	0.00
	26-34	92	390,669	381,164	381,164	2.49	0.00
	35-49	134	727,746	727,746	727,746	0.00	0.00
	50-64	82	886,581	751,019	751,019	18.05	0.00
	65+	37	375,609	511,171	511,171	-26.52	-0.00
Race	**White**	710	2,416,965	2,525,821	2,525,821	-4.31	-0.00
	Black or African American	138	352,646	319,367	319,367	10.42	0.00
	Other	116	264,630	189,053	189,053	39.98	0.00
Hispanicity	**Hispanic or Latino**	175	336,699	390,632	390,632	-13.81	0.00
	Non-Hispanic or Latino	789	2,697,541	2,643,609	2,643,609	2.04	0.00
Gender	**Male**	471	1,466,348	1,460,698	1,460,698	0.39	0.00
	Female	493	1,567,893	1,573,543	1,573,543	-0.36	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.9 2012 NSDUH Slippage Rates: DELAWARE

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		893	765,733	765,733	765,733	0.00	0.00
Quarter	**Quarter 1**	236	190,655	190,655	190,655	0.00	0.00
	Quarter 2	208	191,195	191,195	191,195	-0.00	0.00
	Quarter 3	269	191,712	191,712	191,712	-0.00	0.00
	Quarter 4	180	192,170	192,170	192,170	0.00	0.00
Age Group	**12-17**	306	69,481	68,973	68,973	0.74	0.00
	18-25	244	100,984	102,090	102,090	-1.08	0.00
	26-34	83	103,452	101,033	101,033	2.39	0.00
	35-49	136	169,005	173,393	173,393	-2.53	0.00
	50-64	76	202,506	183,801	183,801	10.18	0.00
	65+	48	120,304	136,443	136,443	-11.83	0.00
Race	**White**	592	549,846	560,218	560,218	-1.85	0.00
	Black or African American	201	157,264	160,461	160,461	-1.99	0.00
	Other	100	58,622	45,054	45,054	30.12	0.00
Hispanicity	**Hispanic or Latino**	133	63,130	57,416	57,416	9.95	0.00
	Non-Hispanic or Latino	760	702,603	708,317	708,317	-0.81	0.00
Gender	**Male**	429	364,315	363,950	363,950	0.10	0.00
	Female	464	401,418	401,783	401,783	-0.09	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.10 2012 NSDUH Slippage Rates: DISTRICT OF COLUMBIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		962	544,627	544,627	544,627	0.00	0.00
Quarter	**Quarter 1**	256	135,282	135,282	135,282	0.00	0.00
	Quarter 2	204	135,824	135,824	135,824	0.00	-0.00
	Quarter 3	242	136,430	136,430	136,430	0.00	0.00
	Quarter 4	260	137,091	137,091	137,091	0.00	0.00
Age Group	**12-17**	329	31,338	31,338	31,338	0.00	0.00
	18-25	341	95,383	95,556	95,556	-0.18	0.00
	26-34	100	124,893	122,971	122,971	1.56	0.00
	35-49	105	117,759	121,246	121,246	-2.88	-0.00
	50-64	52	107,379	104,024	104,024	3.23	0.00
	65+	35	67,875	69,493	69,493	-2.33	0.00
Race	**White**	343	222,189	242,433	242,433	-8.35	-0.00
	Black or African American	507	264,730	264,519	264,519	0.08	-0.00
	Other	112	57,708	37,675	37,675	53.17	0.00
Hispanicity	**Hispanic or Latino**	110	50,163	50,571	50,571	-0.81	0.00
	Non-Hispanic or Latino	852	494,465	494,057	494,057	0.08	0.00
Gender	**Male**	408	251,798	252,737	252,737	-0.37	0.00
	Female	554	292,829	291,890	291,890	0.32	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.11 2012 NSDUH Slippage Rates: FLORIDA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,544	16,382,543	16,382,543	16,382,543	0.00	0.00
Quarter	**Quarter 1**	836	4,074,256	4,074,256	4,074,256	0.00	0.00
	Quarter 2	992	4,088,179	4,088,179	4,088,179	0.00	0.00
	Quarter 3	901	4,102,828	4,102,828	4,102,828	0.00	0.00
	Quarter 4	815	4,117,279	4,117,279	4,117,279	0.00	0.00
Age Group	**12-17**	1,195	1,388,809	1,383,312	1,383,312	0.40	0.00
	18-25	1,211	1,970,708	1,970,724	1,970,724	-0.00	0.00
	26-34	279	2,071,810	2,080,130	2,080,130	-0.40	0.00
	35-49	444	3,703,806	3,680,312	3,680,312	0.64	0.00
	50-64	255	4,279,182	3,823,676	3,823,676	11.91	0.00
	65+	160	2,968,228	3,444,389	3,444,389	-13.82	0.00
Race	**White**	2,427	12,427,353	13,089,877	13,089,877	-5.06	0.00
	Black or African American	750	2,681,193	2,517,863	2,517,863	6.49	0.00
	Other	367	1,273,997	774,803	774,803	64.43	0.00
Hispanicity	**Hispanic or Latino**	1,088	3,761,435	3,686,282	3,686,282	2.04	0.00
	Non-Hispanic or Latino	2,456	12,621,107	12,696,261	12,696,261	-0.59	0.00
Gender	**Male**	1,725	7,864,363	7,869,345	7,869,345	-0.06	0.00
	Female	1,819	8,518,180	8,513,198	8,513,198	0.06	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.12 2012 NSDUH Slippage Rates: GEORGIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		885	8,040,955	8,040,955	8,040,955	-0.00	-0.00
Quarter	**Quarter 1**	190	2,000,123	2,000,123	2,000,123	0.00	0.00
	Quarter 2	261	2,007,469	2,007,469	2,007,469	-0.00	0.00
	Quarter 3	241	2,013,944	2,013,944	2,013,944	0.00	0.00
	Quarter 4	193	2,019,418	2,019,418	2,019,418	-0.00	-0.00
Age Group	**12-17**	284	819,197	828,383	828,383	-1.11	0.00
	18-25	281	1,090,125	1,096,583	1,096,583	-0.59	0.00
	26-34	102	1,197,151	1,173,161	1,173,161	2.04	0.00
	35-49	115	2,031,171	2,029,627	2,029,627	0.08	0.00
	50-64	68	2,009,902	1,805,476	1,805,476	11.32	-0.00
	65+	35	893,409	1,107,725	1,107,725	-19.35	0.00
Race	**White**	519	5,070,473	5,161,502	5,161,502	-1.76	-0.00
	Black or African American	306	2,375,921	2,427,523	2,427,523	-2.13	0.00
	Other	60	594,562	451,930	451,930	31.56	0.00
Hispanicity	**Hispanic or Latino**	76	671,135	645,424	645,424	3.98	-0.00
	Non-Hispanic or Latino	809	7,369,820	7,395,530	7,395,530	-0.35	0.00
Gender	**Male**	410	3,830,159	3,830,159	3,830,159	-0.00	0.00
	Female	475	4,210,796	4,210,796	4,210,796	-0.00	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.13 2012 NSDUH Slippage Rates: HAWAII

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		938	1,130,820	1,130,820	1,130,820	-0.00	0.00
Quarter	**Quarter 1**	163	281,692	281,692	281,692	-0.00	-0.00
	Quarter 2	273	282,365	282,365	282,365	0.00	0.00
	Quarter 3	268	283,046	283,046	283,046	-0.00	0.00
	Quarter 4	234	283,719	283,719	283,719	0.00	0.00
Age Group	**12-17**	283	96,580	96,933	96,933	-0.36	0.00
	18-25	299	140,308	140,267	140,267	0.03	-0.00
	26-34	106	157,536	159,898	159,898	-1.48	-0.00
	35-49	124	251,531	252,207	252,207	-0.27	0.00
	50-64	75	295,019	274,763	274,763	7.37	0.00
	65+	51	189,846	206,753	206,753	-8.18	-0.00
Race	**White**	202	256,177	292,399	292,399	-12.39	0.00
	Black or African American	16	17,755	16,815	16,815	5.59	0.00
	Other	720	856,889	821,607	821,607	4.29	0.00
Hispanicity	**Hispanic or Latino**	144	119,287	90,966	90,966	31.13	0.00
	Non-Hispanic or Latino	794	1,011,533	1,039,854	1,039,854	-2.72	0.00
Gender	**Male**	424	548,412	551,486	551,486	-0.56	0.00
	Female	514	582,408	579,334	579,334	0.53	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.14 2012 NSDUH Slippage Rates: IDAHO

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		921	1,288,271	1,288,271	1,288,271	-0.00	0.00
Quarter	**Quarter 1**	206	320,747	320,747	320,747	-0.00	0.00
	Quarter 2	242	321,662	321,663	321,663	-0.00	0.00
	Quarter 3	252	322,543	322,543	322,543	0.00	0.00
	Quarter 4	221	323,318	323,318	323,318	0.00	0.00
Age Group	**12-17**	346	139,966	139,664	139,664	0.22	0.00
	18-25	258	170,906	173,325	173,325	-1.40	0.00
	26-34	79	188,628	185,025	185,025	1.95	0.00
	35-49	133	281,654	286,017	286,017	-1.53	0.00
	50-64	70	323,116	296,048	296,048	9.14	0.00
	65+	35	184,000	208,192	208,192	-11.62	0.00
Race	**White**	856	1,211,072	1,216,003	1,216,003	-0.41	0.00
	Black or African American	7	9,536	8,479	8,479	12.48	0.00
	Other	58	67,663	63,789	63,789	6.07	0.00
Hispanicity	**Hispanic or Latino**	134	134,045	131,036	131,036	2.30	0.00
	Non-Hispanic or Latino	787	1,154,226	1,157,234	1,157,234	-0.26	0.00
Gender	**Male**	447	626,893	636,632	636,632	-1.53	0.00
	Female	474	661,378	651,639	651,639	1.49	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.15 2012 NSDUH Slippage Rates: ILLINOIS

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,672	10,680,769	10,680,769	10,680,769	0.00	0.00
Quarter	**Quarter 1**	830	2,666,262	2,666,262	2,666,262	0.00	0.00
	Quarter 2	1,050	2,668,545	2,668,545	2,668,545	0.00	0.00
	Quarter 3	954	2,671,399	2,671,399	2,671,399	0.00	0.00
	Quarter 4	838	2,674,563	2,674,563	2,674,563	0.00	0.00
Age Group	**12-17**	1,236	1,057,318	1,051,880	1,051,880	0.52	0.00
	18-25	1,175	1,389,069	1,393,334	1,393,334	-0.31	0.00
	26-34	371	1,587,204	1,581,865	1,581,865	0.34	0.00
	35-49	499	2,556,368	2,546,492	2,546,492	0.39	0.00
	50-64	246	2,597,636	2,472,856	2,472,856	5.05	0.00
	65+	145	1,493,174	1,634,341	1,634,341	-8.64	0.00
Race	**White**	2,607	8,204,855	8,427,752	8,427,752	-2.64	-0.00
	Black or African American	595	1,512,457	1,510,702	1,510,702	0.12	0.00
	Other	470	963,457	742,315	742,315	29.79	0.00
Hispanicity	**Hispanic or Latino**	740	1,599,360	1,580,860	1,580,860	1.17	0.00
	Non-Hispanic or Latino	2,932	9,081,409	9,099,909	9,099,909	-0.20	0.00
Gender	**Male**	1,703	5,137,404	5,173,757	5,173,757	-0.70	0.00
	Female	1,969	5,543,365	5,507,012	5,507,012	0.66	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.16 2012 NSDUH Slippage Rates: INDIANA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		911	5,391,372	5,391,372	5,391,372	0.00	0.00
Quarter	**Quarter 1**	219	1,344,832	1,344,832	1,344,832	-0.00	0.00
	Quarter 2	242	1,346,665	1,346,665	1,346,665	0.00	0.00
	Quarter 3	242	1,348,828	1,348,828	1,348,828	0.00	0.00
	Quarter 4	208	1,351,048	1,351,048	1,351,048	0.00	0.00
Age Group	**12-17**	269	538,211	540,535	540,535	-0.43	0.00
	18-25	329	732,677	731,531	731,531	0.16	0.00
	26-34	88	722,901	736,380	736,380	-1.83	0.00
	35-49	113	1,265,162	1,250,505	1,250,505	1.17	0.00
	50-64	59	1,153,559	1,278,242	1,278,242	-9.75	0.00
	65+	53	978,862	854,179	854,179	14.60	0.00
Race	**White**	716	4,551,281	4,730,440	4,730,440	-3.79	0.00
	Black or African American	141	501,982	472,553	472,553	6.23	0.00
	Other	54	338,109	188,380	188,380	79.48	0.00
Hispanicity	**Hispanic or Latino**	54	289,829	293,430	293,430	-1.23	0.00
	Non-Hispanic or Latino	857	5,101,543	5,097,943	5,097,943	0.07	0.00
Gender	**Male**	425	2,619,075	2,619,075	2,619,075	-0.00	0.00
	Female	486	2,772,297	2,772,297	2,772,297	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.17 2012 NSDUH Slippage Rates: IOWA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		900	2,550,660	2,550,660	2,550,660	-0.00	0.00
Quarter	**Quarter 1**	204	636,274	636,274	636,274	-0.00	0.00
	Quarter 2	246	637,022	637,022	637,022	0.00	0.00
	Quarter 3	248	638,078	638,078	638,078	0.00	0.00
	Quarter 4	202	639,286	639,286	639,286	0.00	0.00
Age Group	**12-17**	314	241,376	241,376	241,376	0.00	0.00
	18-25	282	341,260	347,524	347,524	-1.80	0.00
	26-34	86	372,570	347,794	347,794	7.12	0.00
	35-49	127	540,581	553,408	553,408	-2.32	0.00
	50-64	52	603,902	613,358	613,358	-1.54	0.00
	65+	39	450,972	447,200	447,200	0.84	-0.00
Race	**White**	806	2,360,329	2,389,950	2,389,950	-1.24	0.00
	Black or African American	32	83,687	70,591	70,591	18.55	0.00
	Other	62	106,644	90,119	90,119	18.34	0.00
Hispanicity	**Hispanic or Latino**	52	123,315	114,331	114,331	7.86	0.00
	Non-Hispanic or Latino	848	2,427,345	2,436,329	2,436,329	-0.37	0.00
Gender	**Male**	419	1,255,910	1,255,910	1,255,910	-0.00	0.00
	Female	481	1,294,750	1,294,750	1,294,750	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.18 2012 NSDUH Slippage Rates: KANSAS

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		912	2,336,047	2,336,047	2,336,047	0.00	0.00
Quarter	**Quarter 1**	253	582,418	582,418	582,418	-0.00	0.00
	Quarter 2	211	583,336	583,336	583,336	0.00	0.00
	Quarter 3	245	584,511	584,511	584,511	-0.00	-0.00
	Quarter 4	203	585,782	585,782	585,782	0.00	0.00
Age Group	**12-17**	341	235,155	236,447	236,447	-0.55	0.00
	18-25	264	320,921	322,233	322,233	-0.41	0.00
	26-34	83	337,567	335,039	335,039	0.75	0.00
	35-49	119	515,109	515,034	515,034	0.01	0.00
	50-64	63	571,406	550,930	550,930	3.72	0.00
	65+	42	355,890	376,365	376,365	-5.44	0.00
Race	**White**	768	2,005,386	2,066,667	2,066,667	-2.97	0.00
	Black or African American	54	114,455	132,674	132,674	-13.73	0.00
	Other	90	216,206	136,707	136,707	58.15	0.00
Hispanicity	**Hispanic or Latino**	109	225,691	222,044	222,044	1.64	0.00
	Non-Hispanic or Latino	803	2,110,356	2,114,003	2,114,003	-0.17	-0.00
Gender	**Male**	443	1,132,651	1,142,332	1,142,332	-0.85	0.00
	Female	469	1,203,396	1,193,715	1,193,716	0.81	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.19 2012 NSDUH Slippage Rates: KENTUCKY

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		927	3,607,428	3,607,428	3,607,428	0.00	-0.00
Quarter	**Quarter 1**	203	900,330	900,330	900,330	0.00	0.00
	Quarter 2	267	901,238	901,238	901,238	0.00	0.00
	Quarter 3	215	902,360	902,360	902,360	0.00	0.00
	Quarter 4	242	903,500	903,500	903,500	0.00	-0.00
Age Group	**12-17**	319	343,247	339,442	339,442	1.12	-0.00
	18-25	297	453,476	461,440	461,441	-1.73	-0.00
	26-34	90	527,834	488,287	488,287	8.10	0.00
	35-49	118	852,711	849,738	849,738	0.35	0.00
	50-64	65	945,037	876,308	876,308	7.84	0.00
	65+	38	485,123	592,212	592,213	-18.08	-0.00
Race	**White**	773	3,205,547	3,239,757	3,239,757	-1.06	-0.00
	Black or African American	103	271,780	269,066	269,066	1.01	0.00
	Other	51	130,101	98,605	98,605	31.94	0.00
Hispanicity	**Hispanic or Latino**	41	104,264	96,594	96,594	7.94	0.00
	Non-Hispanic or Latino	886	3,503,164	3,510,834	3,510,834	-0.22	-0.00
Gender	**Male**	454	1,745,937	1,742,383	1,742,383	0.20	-0.00
	Female	473	1,861,490	1,865,044	1,865,044	-0.19	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.20 2012 NSDUH Slippage Rates: LOUISIANA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		901	3,745,460	3,745,460	3,745,460	0.00	0.00
Quarter	**Quarter 1**	200	933,816	933,816	933,816	0.00	0.00
	Quarter 2	253	935,395	935,395	935,395	-0.00	0.00
	Quarter 3	243	937,199	937,199	937,199	0.00	0.00
	Quarter 4	205	939,051	939,051	939,051	0.00	0.00
Age Group	**12-17**	292	367,661	367,661	367,661	0.00	-0.00
	18-25	299	516,355	523,034	523,034	-1.28	0.00
	26-34	86	565,130	558,452	558,452	1.20	0.00
	35-49	117	822,039	837,015	837,015	-1.79	0.00
	50-64	72	981,927	884,266	884,266	11.04	0.00
	65+	35	492,349	575,033	575,033	-14.38	0.00
Race	**White**	532	2,312,694	2,456,732	2,456,732	-5.86	0.00
	Black or African American	303	1,159,687	1,153,099	1,153,099	0.57	0.00
	Other	66	273,080	135,630	135,630	101.34	0.00
Hispanicity	**Hispanic or Latino**	50	230,599	160,281	160,281	43.87	0.00
	Non-Hispanic or Latino	851	3,514,861	3,585,179	3,585,179	-1.96	0.00
Gender	**Male**	399	1,785,505	1,785,505	1,785,505	0.00	0.00
	Female	502	1,959,955	1,959,955	1,959,955	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.21 2012 NSDUH Slippage Rates: MAINE

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		938	1,145,565	1,145,565	1,145,565	-0.00	0.00
Quarter	**Quarter 1**	237	286,125	286,125	286,125	-0.00	0.00
	Quarter 2	228	286,317	286,317	286,317	0.00	0.00
	Quarter 3	253	286,508	286,508	286,508	-0.00	0.00
	Quarter 4	220	286,615	286,615	286,615	0.00	0.00
Age Group	**12-17**	307	96,759	95,666	95,666	1.14	0.00
	18-25	320	125,098	129,416	129,416	-3.34	0.00
	26-34	81	136,192	131,286	131,286	3.74	0.00
	35-49	112	257,425	259,106	259,106	-0.65	0.00
	50-64	65	287,923	310,862	310,862	-7.38	0.00
	65+	53	242,169	219,230	219,230	10.46	0.00
Race	**White**	861	1,092,309	1,099,016	1,099,016	-0.61	0.00
	Black or African American	21	18,564	25,248	12,610	47.21	100.21
	Other	56	34,693	21,302	33,939	2.22	-37.24
Hispanicity	**Hispanic or Latino**	28	16,957	13,687	13,687	23.89	0.00
	Non-Hispanic or Latino	910	1,128,609	1,131,878	1,131,878	-0.29	0.00
Gender	**Male**	456	560,379	555,720	555,720	0.84	0.00
	Female	482	585,186	589,846	589,846	-0.79	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.22 2012 NSDUH Slippage Rates: MARYLAND

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		874	4,905,827	4,905,827	4,905,827	-0.00	0.00
Quarter	**Quarter 1**	228	1,222,175	1,222,175	1,222,175	0.00	0.00
	Quarter 2	220	1,225,004	1,225,004	1,225,004	-0.00	0.00
	Quarter 3	217	1,227,919	1,227,919	1,227,919	-0.00	0.00
	Quarter 4	209	1,230,729	1,230,729	1,230,729	0.00	0.00
Age Group	**12-17**	281	456,705	458,368	458,368	-0.36	0.00
	18-25	295	605,664	631,975	631,975	-4.16	0.00
	26-34	82	733,808	698,875	698,875	5.00	0.00
	35-49	123	1,160,444	1,196,316	1,196,316	-3.00	0.00
	50-64	53	1,159,101	1,180,907	1,180,907	-1.85	0.00
	65+	40	790,105	739,386	739,386	6.86	0.00
Race	**White**	517	2,860,779	3,047,799	3,047,799	-6.14	0.00
	Black or African American	246	1,424,232	1,435,721	1,435,721	-0.80	0.00
	Other	111	620,816	422,307	422,307	47.01	0.00
Hispanicity	**Hispanic or Latino**	65	393,167	390,124	390,124	0.78	0.00
	Non-Hispanic or Latino	809	4,512,660	4,515,703	4,515,703	-0.07	0.00
Gender	**Male**	415	2,331,715	2,331,715	2,331,715	-0.00	0.00
	Female	459	2,574,111	2,574,111	2,574,111	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.23 2012 NSDUH Slippage Rates: MASSACHUSETTS

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		955	5,661,530	5,661,530	5,661,530	-0.00	0.00
Quarter	**Quarter 1**	192	1,410,999	1,410,999	1,410,999	-0.00	0.00
	Quarter 2	211	1,413,753	1,413,753	1,413,753	0.00	0.00
	Quarter 3	248	1,416,840	1,416,840	1,416,840	0.00	0.00
	Quarter 4	304	1,419,938	1,419,938	1,419,938	0.00	-0.00
Age Group	**12-17**	310	495,413	493,395	493,395	0.41	0.00
	18-25	307	758,927	772,360	772,360	-1.74	-0.00
	26-34	93	791,512	786,066	786,066	0.69	-0.00
	35-49	135	1,312,441	1,339,344	1,339,344	-2.01	0.00
	50-64	70	1,519,568	1,350,049	1,350,049	12.56	0.00
	65+	40	783,671	920,318	920,318	-14.85	-0.00
Race	**White**	771	4,669,770	4,792,328	4,792,328	-2.56	0.00
	Black or African American	80	437,706	409,776	422,413	3.62	-2.99
	Other	104	554,054	459,427	446,790	24.01	2.83
Hispanicity	**Hispanic or Latino**	151	553,293	517,569	517,569	6.90	0.00
	Non-Hispanic or Latino	804	5,108,236	5,143,961	5,143,961	-0.69	-0.00
Gender	**Male**	477	2,714,719	2,713,214	2,713,214	0.06	-0.00
	Female	478	2,946,811	2,948,316	2,948,316	-0.05	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.24 2012 NSDUH Slippage Rates: MICHIGAN

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,655	8,319,227	8,319,227	8,319,227	0.00	0.00
Quarter	**Quarter 1**	850	2,076,801	2,076,801	2,076,801	0.00	0.00
	Quarter 2	1,101	2,078,568	2,078,568	2,078,568	0.00	-0.00
	Quarter 3	955	2,080,800	2,080,800	2,080,800	0.00	0.00
	Quarter 4	749	2,083,057	2,083,057	2,083,057	0.00	0.00
Age Group	**12-17**	1,175	808,270	809,401	809,401	-0.14	0.00
	18-25	1,230	1,109,543	1,101,787	1,101,787	0.70	0.00
	26-34	319	1,020,053	1,040,667	1,040,667	-1.98	0.00
	35-49	488	1,877,606	1,895,536	1,895,536	-0.95	0.00
	50-64	276	2,203,039	2,065,859	2,065,859	6.64	-0.00
	65+	167	1,300,716	1,405,976	1,405,976	-7.49	-0.00
Race	**White**	2,825	6,672,978	6,775,719	6,775,719	-1.52	-0.00
	Black or African American	556	1,142,446	1,132,805	1,132,805	0.85	-0.00
	Other	274	503,802	410,703	410,703	22.67	0.00
Hispanicity	**Hispanic or Latino**	230	319,193	332,530	332,530	-4.01	0.00
	Non-Hispanic or Latino	3,425	8,000,033	7,986,697	7,986,697	0.17	0.00
Gender	**Male**	1,849	4,028,991	4,029,940	4,029,940	-0.02	0.00
	Female	1,806	4,290,236	4,289,287	4,289,287	0.02	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.25 2012 NSDUH Slippage Rates: MINNESOTA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		902	4,470,678	4,470,679	4,470,679	-0.00	0.00
Quarter	**Quarter 1**	263	1,114,212	1,114,212	1,114,212	0.00	0.00
	Quarter 2	210	1,116,468	1,116,468	1,116,468	0.00	0.00
	Quarter 3	202	1,118,861	1,118,861	1,118,861	-0.00	0.00
	Quarter 4	227	1,121,138	1,121,138	1,121,138	0.00	0.00
Age Group	**12-17**	325	425,742	424,357	424,357	0.33	0.00
	18-25	263	553,684	571,203	571,203	-3.07	0.00
	26-34	106	670,384	658,358	658,358	1.83	0.00
	35-49	122	1,035,858	1,035,209	1,035,209	0.06	0.00
	50-64	45	911,798	1,081,348	1,081,348	-15.68	0.00
	65+	41	873,213	700,204	700,204	24.71	-0.00
Race	**White**	747	3,928,755	3,942,889	3,942,889	-0.36	0.00
	Black or African American	52	234,645	218,218	218,218	7.53	-0.00
	Other	103	307,279	309,572	309,572	-0.74	0.00
Hispanicity	**Hispanic or Latino**	69	194,638	185,215	185,215	5.09	0.00
	Non-Hispanic or Latino	833	4,276,040	4,285,464	4,285,464	-0.22	0.00
Gender	**Male**	456	2,205,477	2,205,477	2,205,477	0.00	0.00
	Female	446	2,265,202	2,265,202	2,265,202	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.26 2012 NSDUH Slippage Rates: MISSISSIPPI

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		901	2,419,811	2,419,811	2,419,811	0.00	-0.00
Quarter	**Quarter 1**	227	603,821	603,821	603,821	0.00	-0.00
	Quarter 2	204	604,547	604,547	604,547	0.00	0.00
	Quarter 3	274	605,332	605,332	605,332	0.00	-0.00
	Quarter 4	196	606,111	606,111	606,111	0.00	-0.00
Age Group	**12-17**	310	245,489	248,208	248,208	-1.10	-0.00
	18-25	297	335,636	336,270	336,270	-0.19	0.00
	26-34	78	352,700	333,320	333,320	5.81	0.00
	35-49	115	515,908	547,591	547,591	-5.79	0.00
	50-64	68	640,242	564,294	564,294	13.46	0.00
	65+	33	329,836	390,127	390,127	-15.45	-0.00
Race	**White**	450	1,466,616	1,493,746	1,493,746	-1.82	-0.00
	Black or African American	397	860,236	869,000	869,000	-1.01	0.00
	Other	54	92,959	57,064	57,065	62.90	-0.00
Hispanicity	**Hispanic or Latino**	35	65,411	59,333	59,333	10.24	0.00
	Non-Hispanic or Latino	866	2,354,400	2,360,478	2,360,478	-0.26	-0.00
Gender	**Male**	408	1,140,730	1,143,501	1,143,501	-0.24	0.00
	Female	493	1,279,081	1,276,310	1,276,310	0.22	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.27 2012 NSDUH Slippage Rates: MISSOURI

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		915	4,985,565	4,985,565	4,985,565	-0.00	0.00
Quarter	**Quarter 1**	190	1,244,091	1,244,091	1,244,091	-0.00	0.00
	Quarter 2	286	1,245,468	1,245,468	1,245,468	0.00	0.00
	Quarter 3	227	1,247,132	1,247,132	1,247,132	0.00	0.00
	Quarter 4	212	1,248,874	1,248,874	1,248,874	0.00	0.00
Age Group	**12-17**	313	480,051	474,059	474,059	1.26	0.00
	18-25	286	643,657	654,819	654,819	-1.70	0.00
	26-34	99	688,579	692,752	692,752	-0.60	0.00
	35-49	114	1,117,774	1,117,411	1,117,411	0.03	0.00
	50-64	67	1,361,391	1,199,689	1,199,689	13.48	0.00
	65+	36	694,113	846,836	846,836	-18.03	-0.00
Race	**White**	755	4,166,191	4,238,678	4,238,678	-1.71	0.00
	Black or African American	100	579,156	550,971	550,971	5.12	0.00
	Other	60	240,219	195,916	195,916	22.61	0.00
Hispanicity	**Hispanic or Latino**	38	150,798	161,457	161,457	-6.60	0.00
	Non-Hispanic or Latino	877	4,834,767	4,824,108	4,824,108	0.22	0.00
Gender	**Male**	437	2,422,297	2,405,500	2,405,500	0.70	0.00
	Female	478	2,563,268	2,580,065	2,580,065	-0.65	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.28 2012 NSDUH Slippage Rates: MONTANA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		876	842,009	842,009	842,009	0.00	0.00
Quarter	**Quarter 1**	267	209,828	209,828	209,828	0.00	0.00
	Quarter 2	181	210,273	210,274	210,274	-0.00	0.00
	Quarter 3	200	210,739	210,739	210,739	-0.00	0.00
	Quarter 4	228	211,169	211,169	211,169	-0.00	0.00
Age Group	**12-17**	313	73,150	73,775	73,775	-0.85	0.00
	18-25	280	107,786	107,843	107,843	-0.05	0.00
	26-34	78	108,340	110,864	110,864	-2.28	0.00
	35-49	110	175,600	173,862	173,862	1.00	0.00
	50-64	62	248,252	222,066	222,066	11.79	0.00
	65+	33	128,881	153,600	153,600	-16.09	0.00
Race	**White**	774	771,013	766,472	766,472	0.59	0.00
	Black or African American	5	2,451	4,214	4,214	-41.83	-0.00
	Other	97	68,545	71,323	71,323	-3.90	0.00
Hispanicity	**Hispanic or Latino**	39	26,810	23,076	23,076	16.18	0.00
	Non-Hispanic or Latino	837	815,199	818,933	818,933	-0.46	0.00
Gender	**Male**	449	419,133	419,133	419,133	0.00	0.00
	Female	427	422,876	422,876	422,876	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.29 2012 NSDUH Slippage Rates: NEBRASKA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		940	1,511,302	1,511,302	1,511,302	0.00	0.00
Quarter	**Quarter 1**	227	376,592	376,592	376,592	-0.00	0.00
	Quarter 2	225	377,349	377,349	377,349	-0.00	0.00
	Quarter 3	252	378,224	378,224	378,224	-0.00	0.00
	Quarter 4	236	379,137	379,137	379,137	0.00	0.00
Age Group	**12-17**	277	146,413	147,378	147,378	-0.66	0.00
	18-25	365	209,677	205,771	205,771	1.90	0.00
	26-34	88	219,377	222,316	222,316	-1.32	0.00
	35-49	114	332,700	334,018	334,018	-0.39	0.00
	50-64	61	402,902	356,051	356,051	13.16	0.00
	65+	35	200,233	245,767	245,767	-18.53	0.00
Race	**White**	841	1,351,635	1,376,864	1,376,864	-1.83	0.00
	Black or African American	32	67,187	65,811	65,811	2.09	0.00
	Other	67	92,480	68,628	68,628	34.76	0.00
Hispanicity	**Hispanic or Latino**	106	126,869	124,861	124,861	1.61	0.00
	Non-Hispanic or Latino	834	1,384,433	1,386,441	1,386,441	-0.14	0.00
Gender	**Male**	472	743,166	743,167	743,167	-0.00	0.00
	Female	468	768,136	768,136	768,136	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.30 2012 NSDUH Slippage Rates: NEVADA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		903	2,278,656	2,278,656	2,278,656	0.00	0.00
Quarter	**Quarter 1**	233	566,215	566,215	566,215	-0.00	0.00
	Quarter 2	219	568,619	568,619	568,619	0.00	0.00
	Quarter 3	249	570,896	570,896	570,896	0.00	0.00
	Quarter 4	202	572,927	572,927	572,927	-0.00	0.00
Age Group	**12-17**	293	223,823	220,899	220,899	1.32	0.00
	18-25	281	277,256	284,532	284,532	-2.56	0.00
	26-34	124	345,967	343,051	343,051	0.85	0.00
	35-49	113	551,462	559,210	559,210	-1.39	0.00
	50-64	52	487,196	515,202	515,202	-5.44	0.00
	65+	40	392,952	355,762	355,762	10.45	0.00
Race	**White**	627	1,637,063	1,777,070	1,777,070	-7.88	0.00
	Black or African American	78	193,269	192,970	192,970	0.16	0.00
	Other	198	448,324	308,617	308,617	45.27	0.00
Hispanicity	**Hispanic or Latino**	366	576,836	564,568	564,568	2.17	0.00
	Non-Hispanic or Latino	537	1,701,820	1,714,089	1,714,089	-0.72	0.00
Gender	**Male**	433	1,135,354	1,134,734	1,134,734	0.05	0.00
	Female	470	1,143,302	1,143,923	1,143,923	-0.05	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.31 2012 NSDUH Slippage Rates: NEW HAMPSHIRE

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		950	1,133,661	1,133,661	1,133,661	-0.00	0.00
Quarter	**Quarter 1**	211	282,814	282,814	282,814	-0.00	0.00
	Quarter 2	229	283,257	283,257	283,257	0.00	0.00
	Quarter 3	264	283,655	283,655	283,655	-0.00	0.00
	Quarter 4	246	283,935	283,935	283,935	0.00	0.00
Age Group	**12-17**	304	101,749	102,103	102,103	-0.35	0.00
	18-25	323	142,236	139,482	139,482	1.97	0.00
	26-34	62	124,708	131,890	131,890	-5.45	-0.00
	35-49	141	277,735	270,957	270,957	2.50	0.00
	50-64	82	328,265	302,808	302,808	8.41	0.00
	65+	38	158,968	186,422	186,422	-14.73	0.00
Race	**White**	876	1,063,444	1,077,004	1,077,004	-1.26	0.00
	Black or African American	14	12,943	14,536	14,536	-10.96	0.00
	Other	60	57,274	42,122	42,122	35.97	0.00
Hispanicity	**Hispanic or Latino**	54	31,477	29,943	29,943	5.12	0.00
	Non-Hispanic or Latino	896	1,102,185	1,103,718	1,103,718	-0.14	0.00
Gender	**Male**	469	556,164	556,164	556,164	0.00	0.00
	Female	481	577,498	577,498	577,498	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.32 2012 NSDUH Slippage Rates: NEW JERSEY

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		898	7,440,994	7,440,994	7,440,994	0.00	0.00
Quarter	**Quarter 1**	157	1,855,931	1,855,931	1,855,931	0.00	0.00
	Quarter 2	315	1,858,445	1,858,445	1,858,445	-0.00	0.00
	Quarter 3	220	1,861,594	1,861,594	1,861,594	0.00	0.00
	Quarter 4	206	1,865,024	1,865,024	1,865,024	-0.00	0.00
Age Group	**12-17**	291	712,730	708,659	708,659	0.57	0.00
	18-25	287	871,835	881,583	881,583	-1.11	0.00
	26-34	84	980,064	1,004,265	1,004,265	-2.41	0.00
	35-49	137	1,884,319	1,854,441	1,854,441	1.61	0.00
	50-64	51	1,553,322	1,779,911	1,779,911	-12.73	0.00
	65+	48	1,438,724	1,212,136	1,212,136	18.69	0.00
Race	**White**	534	4,975,343	5,569,957	5,569,957	-10.68	0.00
	Black or African American	177	1,174,144	1,043,236	1,043,236	12.55	0.00
	Other	187	1,291,507	827,802	827,802	56.02	0.00
Hispanicity	**Hispanic or Latino**	221	1,390,786	1,296,529	1,296,529	7.27	0.00
	Non-Hispanic or Latino	677	6,050,208	6,144,466	6,144,466	-1.53	0.00
Gender	**Male**	442	3,579,835	3,579,260	3,579,260	0.02	0.00
	Female	456	3,861,159	3,861,734	3,861,734	-0.01	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.33 2012 NSDUH Slippage Rates: NEW MEXICO

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		879	1,702,667	1,702,667	1,702,667	-0.00	0.00
Quarter	**Quarter 1**	213	424,965	424,965	424,965	0.00	0.00
	Quarter 2	240	425,418	425,419	425,419	-0.00	0.00
	Quarter 3	216	425,917	425,917	425,917	0.00	0.00
	Quarter 4	210	426,367	426,367	426,367	-0.00	0.00
Age Group	**12-17**	289	168,547	168,839	168,839	-0.17	0.00
	18-25	303	231,620	226,708	226,708	2.17	0.00
	26-34	79	224,497	237,352	237,352	-5.42	0.00
	35-49	112	365,656	368,268	368,268	-0.71	0.00
	50-64	61	459,537	411,923	411,923	11.56	0.00
	65+	35	252,809	289,577	289,577	-12.70	0.00
Race	**White**	687	1,426,380	1,434,076	1,434,076	-0.54	0.00
	Black or African American	17	39,764	38,542	38,542	3.17	0.00
	Other	175	236,523	230,049	230,049	2.81	0.00
Hispanicity	**Hispanic or Latino**	466	772,919	762,809	762,809	1.33	0.00
	Non-Hispanic or Latino	413	929,748	939,858	939,858	-1.08	0.00
Gender	**Male**	412	827,048	827,048	827,048	-0.00	0.00
	Female	467	875,619	875,620	875,620	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.34 2012 NSDUH Slippage Rates: NEW YORK

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,680	16,532,006	16,532,006	16,532,006	-0.00	-0.00
Quarter	**Quarter 1**	903	4,124,895	4,124,896	4,124,896	-0.00	0.00
	Quarter 2	982	4,129,394	4,129,394	4,129,394	-0.00	0.00
	Quarter 3	881	4,135,432	4,135,432	4,135,432	-0.00	-0.00
	Quarter 4	914	4,142,284	4,142,284	4,142,284	-0.00	0.00
Age Group	**12-17**	1,193	1,468,023	1,466,519	1,466,519	0.10	-0.00
	18-25	1,253	2,252,607	2,246,785	2,246,785	0.26	-0.00
	26-34	338	2,386,731	2,430,252	2,430,252	-1.79	0.00
	35-49	509	4,042,348	3,898,670	3,898,670	3.69	-0.00
	50-64	234	3,822,619	3,826,575	3,826,575	-0.10	-0.00
	65+	153	2,559,678	2,663,205	2,663,205	-3.89	0.00
Race	**White**	2,446	10,939,101	11,919,999	11,919,999	-8.23	-0.00
	Black or African American	624	2,836,979	2,784,179	2,784,180	1.90	-0.00
	Other	610	2,755,927	1,827,827	1,827,827	50.78	-0.00
Hispanicity	**Hispanic or Latino**	823	2,895,937	2,843,870	2,843,870	1.83	-0.00
	Non-Hispanic or Latino	2,857	13,636,069	13,688,136	13,688,136	-0.38	-0.00
Gender	**Male**	1,766	7,897,578	7,901,430	7,901,430	-0.05	0.00
	Female	1,914	8,634,428	8,630,576	8,630,576	0.04	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.35 2012 NSDUH Slippage Rates: NORTH CAROLINA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		917	8,007,328	8,007,328	8,007,328	0.00	0.00
Quarter	**Quarter 1**	190	1,992,670	1,992,670	1,992,670	-0.00	0.00
	Quarter 2	269	1,998,901	1,998,901	1,998,901	0.00	0.00
	Quarter 3	243	2,005,037	2,005,037	2,005,037	0.00	0.00
	Quarter 4	215	2,010,721	2,010,721	2,010,721	-0.00	0.00
Age Group	**12-17**	298	760,601	760,601	760,601	0.00	0.00
	18-25	334	1,023,840	1,033,454	1,033,454	-0.93	0.00
	26-34	65	1,076,555	1,086,873	1,086,873	-0.95	0.00
	35-49	126	1,947,225	1,948,167	1,948,167	-0.05	0.00
	50-64	46	1,549,677	1,871,994	1,871,994	-17.22	0.00
	65+	48	1,649,429	1,306,240	1,306,240	26.27	0.00
Race	**White**	605	5,725,330	5,862,050	5,862,050	-2.33	0.00
	Black or African American	214	1,744,682	1,710,790	1,710,791	1.98	-0.00
	Other	98	537,316	434,487	434,487	23.67	0.00
Hispanicity	**Hispanic or Latino**	84	561,780	592,610	592,610	-5.20	0.00
	Non-Hispanic or Latino	833	7,445,548	7,414,718	7,414,718	0.42	0.00
Gender	**Male**	438	3,806,696	3,804,112	3,804,112	0.07	0.00
	Female	479	4,200,632	4,203,216	4,203,216	-0.06	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.36 2012 NSDUH Slippage Rates: NORTH DAKOTA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		895	577,526	577,526	577,526	-0.00	-0.00
Quarter	**Quarter 1**	202	143,198	143,198	143,198	-0.00	0.00
	Quarter 2	230	143,924	143,924	143,924	-0.00	0.00
	Quarter 3	273	144,752	144,752	144,752	0.00	-0.00
	Quarter 4	190	145,652	145,652	145,652	0.00	0.00
Age Group	**12-17**	307	48,628	48,911	48,912	-0.58	-0.00
	18-25	269	92,648	93,645	93,645	-1.06	0.00
	26-34	91	86,172	84,751	84,751	1.68	0.00
	35-49	116	116,190	117,448	117,448	-1.07	-0.00
	50-64	70	150,996	137,722	137,722	9.64	0.00
	65+	42	82,891	95,049	95,049	-12.79	0.00
Race	**White**	784	528,755	528,175	528,175	0.11	-0.00
	Black or African American	18	7,191	7,191	7,191	-0.00	0.00
	Other	93	41,580	42,160	42,160	-1.38	0.00
Hispanicity	**Hispanic or Latino**	30	11,919	12,285	12,285	-2.98	-0.00
	Non-Hispanic or Latino	865	565,607	565,241	565,241	0.06	0.00
Gender	**Male**	452	291,063	291,063	291,063	-0.00	-0.00
	Female	443	286,463	286,463	286,463	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.37 2012 NSDUH Slippage Rates: OHIO

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,687	9,638,652	9,638,652	9,638,652	0.00	0.00
Quarter	**Quarter 1**	822	2,406,762	2,406,762	2,406,762	-0.00	0.00
	Quarter 2	937	2,408,346	2,408,346	2,408,346	0.00	0.00
	Quarter 3	994	2,410,585	2,410,585	2,410,585	0.00	0.00
	Quarter 4	934	2,412,958	2,412,958	2,412,958	-0.00	0.00
Age Group	**12-17**	1,291	922,961	926,791	926,791	-0.41	0.00
	18-25	1,138	1,219,903	1,232,694	1,232,694	-1.04	0.00
	26-34	332	1,287,601	1,261,992	1,261,992	2.03	0.00
	35-49	486	2,184,399	2,193,386	2,193,386	-0.41	0.00
	50-64	269	2,490,936	2,386,064	2,386,064	4.40	0.00
	65+	171	1,532,852	1,637,724	1,637,724	-6.40	-0.00
Race	**White**	2,938	8,039,924	8,167,761	8,167,761	-1.57	0.00
	Black or African American	499	1,137,855	1,128,931	1,128,931	0.79	0.00
	Other	250	460,873	341,960	341,960	34.77	0.00
Hispanicity	**Hispanic or Latino**	164	282,356	271,903	271,903	3.84	0.00
	Non-Hispanic or Latino	3,523	9,356,296	9,366,749	9,366,749	-0.11	0.00
Gender	**Male**	1,773	4,652,147	4,650,007	4,650,007	0.05	0.00
	Female	1,914	4,986,505	4,988,645	4,988,645	-0.04	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.38 2012 NSDUH Slippage Rates: OKLAHOMA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		908	3,099,247	3,099,247	3,099,247	0.00	-0.00
Quarter	**Quarter 1**	218	771,867	771,867	771,867	0.00	-0.00
	Quarter 2	266	773,765	773,765	773,765	0.00	-0.00
	Quarter 3	225	775,780	775,780	775,780	0.00	0.00
	Quarter 4	199	777,835	777,835	777,835	-0.00	0.00
Age Group	**12-17**	301	304,249	305,458	305,458	-0.40	-0.00
	18-25	296	425,949	424,952	424,952	0.23	-0.00
	26-34	98	423,443	448,396	448,396	-5.56	0.00
	35-49	102	706,838	683,977	683,977	3.34	0.00
	50-64	78	880,387	719,615	719,615	22.34	0.00
	65+	33	358,381	516,849	516,849	-30.66	0.00
Race	**White**	630	2,344,690	2,400,995	2,400,995	-2.35	0.00
	Black or African American	67	225,531	220,347	220,347	2.35	-0.00
	Other	211	529,026	477,906	477,906	10.70	0.00
Hispanicity	**Hispanic or Latino**	133	281,088	247,100	247,101	13.75	-0.00
	Non-Hispanic or Latino	775	2,818,159	2,852,147	2,852,147	-1.19	0.00
Gender	**Male**	427	1,504,592	1,504,592	1,504,592	0.00	-0.00
	Female	481	1,594,655	1,594,655	1,594,655	0.00	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.39 2012 NSDUH Slippage Rates: OREGON

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		923	3,293,097	3,293,097	3,293,097	-0.00	0.00
Quarter	**Quarter 1**	227	820,199	820,199	820,199	-0.00	-0.00
	Quarter 2	209	822,306	822,306	822,306	-0.00	0.00
	Quarter 3	251	824,382	824,382	824,382	-0.00	0.00
	Quarter 4	236	826,210	826,210	826,210	0.00	0.00
Age Group	**12-17**	270	294,453	292,395	292,395	0.70	-0.00
	18-25	314	404,332	409,756	409,756	-1.32	0.00
	26-34	80	467,857	474,426	474,426	-1.38	-0.00
	35-49	132	748,652	744,523	744,523	0.55	0.00
	50-64	80	834,706	800,891	800,891	4.22	0.00
	65+	47	543,097	571,107	571,107	-4.90	0.00
Race	**White**	768	2,819,380	2,938,699	2,938,699	-4.06	-0.00
	Black or African American	29	66,146	60,572	60,572	9.20	-0.00
	Other	126	407,572	293,826	293,826	38.71	0.00
Hispanicity	**Hispanic or Latino**	140	344,975	341,769	341,769	0.94	0.00
	Non-Hispanic or Latino	783	2,948,122	2,951,328	2,951,328	-0.11	0.00
Gender	**Male**	430	1,597,893	1,612,236	1,612,236	-0.89	0.00
	Female	493	1,695,204	1,680,861	1,680,861	0.85	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.40 2012 NSDUH Slippage Rates: PENNSYLVANIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,580	10,790,033	10,790,033	10,790,033	0.00	0.00
Quarter	**Quarter 1**	773	2,693,509	2,693,509	2,693,509	0.00	0.00
	Quarter 2	1,059	2,695,530	2,695,530	2,695,530	0.00	0.00
	Quarter 3	1,013	2,698,687	2,698,687	2,698,687	0.00	0.00
	Quarter 4	735	2,702,308	2,702,308	2,702,308	-0.00	0.00
Age Group	**12-17**	1,166	956,763	958,552	958,552	-0.19	0.00
	18-25	1,203	1,387,856	1,404,841	1,404,841	-1.21	0.00
	26-34	296	1,374,492	1,377,020	1,377,020	-0.18	0.00
	35-49	466	2,425,111	2,412,498	2,412,498	0.52	0.00
	50-64	273	2,809,570	2,670,080	2,670,080	5.22	0.00
	65+	176	1,836,242	1,967,043	1,967,043	-6.65	-0.00
Race	**White**	2,865	8,984,408	9,158,533	9,158,533	-1.90	-0.00
	Black or African American	423	1,123,749	1,137,744	1,137,744	-1.23	-0.00
	Other	292	681,876	493,756	493,756	38.10	0.00
Hispanicity	**Hispanic or Latino**	322	632,984	573,713	573,713	10.33	0.00
	Non-Hispanic or Latino	3,258	10,157,049	10,216,320	10,216,320	-0.58	0.00
Gender	**Male**	1,699	5,207,979	5,194,845	5,194,845	0.25	0.00
	Female	1,881	5,582,054	5,595,188	5,595,188	-0.23	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.41 2012 NSDUH Slippage Rates: RHODE ISLAND

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		923	895,345	895,345	895,345	0.00	0.00
Quarter	**Quarter 1**	250	223,586	223,586	223,586	0.00	0.00
	Quarter 2	190	223,711	223,711	223,711	0.00	0.00
	Quarter 3	287	223,910	223,910	223,910	0.00	0.00
	Quarter 4	196	224,138	224,138	224,138	0.00	0.00
Age Group	**12-17**	279	79,046	77,245	77,245	2.33	0.00
	18-25	323	130,321	132,691	132,691	-1.79	0.00
	26-34	86	119,688	114,691	114,691	4.36	0.00
	35-49	126	197,598	202,372	202,372	-2.36	0.00
	50-64	71	236,382	216,834	216,834	9.02	0.00
	65+	38	132,310	151,511	151,511	-12.67	-0.00
Race	**White**	719	743,983	780,516	780,516	-4.68	0.00
	Black or African American	65	62,697	60,504	60,504	3.62	-0.00
	Other	139	88,664	54,325	54,325	63.21	0.00
Hispanicity	**Hispanic or Latino**	152	114,157	105,310	105,310	8.40	0.00
	Non-Hispanic or Latino	771	781,188	790,035	790,035	-1.12	0.00
Gender	**Male**	442	427,594	427,594	427,594	0.00	0.00
	Female	481	467,752	467,752	467,752	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.42 2012 NSDUH Slippage Rates: SOUTH CAROLINA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		938	3,900,041	3,900,041	3,900,041	0.00	0.00
Quarter	**Quarter 1**	210	970,470	970,470	970,470	0.00	0.00
	Quarter 2	239	973,808	973,808	973,808	0.00	0.00
	Quarter 3	242	976,619	976,619	976,619	0.00	0.00
	Quarter 4	247	979,144	979,144	979,144	0.00	0.00
Age Group	**12-17**	317	358,471	358,471	358,471	0.00	0.00
	18-25	291	510,740	515,765	515,765	-0.97	0.00
	26-34	89	517,042	520,199	520,199	-0.61	0.00
	35-49	147	900,107	891,926	891,926	0.92	0.00
	50-64	64	1,093,432	936,210	936,210	16.79	0.00
	65+	30	520,249	677,471	677,471	-23.21	-0.00
Race	**White**	543	2,658,359	2,727,078	2,727,078	-2.52	0.00
	Black or African American	332	1,047,509	1,049,920	1,049,920	-0.23	0.00
	Other	63	194,173	123,043	123,043	57.81	0.00
Hispanicity	**Hispanic or Latino**	65	196,970	177,230	177,230	11.14	0.00
	Non-Hispanic or Latino	873	3,703,071	3,722,811	3,722,811	-0.53	0.00
Gender	**Male**	447	1,851,304	1,851,304	1,851,304	0.00	0.00
	Female	491	2,048,737	2,048,737	2,048,737	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.43 2012 NSDUH Slippage Rates: SOUTH DAKOTA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		878	676,283	676,283	676,283	0.00	0.00
Quarter	**Quarter 1**	181	168,181	168,181	168,181	0.00	0.00
	Quarter 2	271	168,753	168,753	168,753	0.00	0.00
	Quarter 3	202	169,372	169,372	169,372	0.00	-0.00
	Quarter 4	224	169,978	169,978	169,978	0.00	0.00
Age Group	**12-17**	266	64,784	64,543	64,543	0.37	0.00
	18-25	297	90,967	91,525	91,525	-0.61	-0.00
	26-34	81	94,548	95,625	95,625	-1.13	0.00
	35-49	105	144,751	143,357	143,357	0.97	0.00
	50-64	77	164,483	165,211	165,211	-0.44	-0.00
	65+	52	116,751	116,023	116,023	0.63	0.00
Race	**White**	747	597,800	596,645	596,645	0.19	0.00
	Black or African American	13	9,881	10,200	10,200	-3.13	0.00
	Other	118	68,602	69,439	69,439	-1.20	0.00
Hispanicity	**Hispanic or Latino**	23	11,086	17,985	17,985	-38.36	0.00
	Non-Hispanic or Latino	855	665,197	658,298	658,298	1.05	-0.00
Gender	**Male**	425	336,855	335,247	335,247	0.48	0.00
	Female	453	339,428	341,036	341,036	-0.47	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.44 2012 NSDUH Slippage Rates: TENNESSEE

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		927	5,363,074	5,363,074	5,363,074	0.00	-0.00
Quarter	**Quarter 1**	209	1,335,505	1,335,505	1,335,505	0.00	-0.00
	Quarter 2	264	1,338,969	1,338,969	1,338,969	0.00	-0.00
	Quarter 3	231	1,342,571	1,342,571	1,342,571	0.00	-0.00
	Quarter 4	223	1,346,029	1,346,029	1,346,029	-0.00	-0.00
Age Group	**12-17**	261	504,426	505,108	505,108	-0.14	-0.00
	18-25	348	689,362	688,253	688,253	0.16	0.00
	26-34	80	723,401	736,176	736,176	-1.74	0.00
	35-49	130	1,291,763	1,267,313	1,267,313	1.93	-0.00
	50-64	69	1,416,591	1,276,804	1,276,804	10.95	-0.00
	65+	39	737,531	889,419	889,419	-17.08	-0.00
Race	**White**	675	4,298,431	4,319,445	4,319,446	-0.49	-0.00
	Black or African American	199	859,193	868,421	868,421	-1.06	0.00
	Other	53	205,449	175,207	175,207	17.26	0.00
Hispanicity	**Hispanic or Latino**	49	220,724	220,076	220,076	0.29	0.00
	Non-Hispanic or Latino	878	5,142,350	5,142,998	5,142,998	-0.01	-0.00
Gender	**Male**	439	2,571,530	2,571,530	2,571,530	0.00	0.00
	Female	488	2,791,544	2,791,544	2,791,544	0.00	-0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.45 2012 NSDUH Slippage Rates: TEXAS

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		3,625	20,852,844	20,852,844	20,852,844	-0.00	0.00
Quarter	**Quarter 1**	894	5,177,708	5,177,708	5,177,708	0.00	0.00
	Quarter 2	942	5,201,200	5,201,200	5,201,200	-0.00	0.00
	Quarter 3	876	5,225,110	5,225,110	5,225,110	0.00	0.00
	Quarter 4	913	5,248,826	5,248,826	5,248,826	-0.00	0.00
Age Group	**12-17**	1,244	2,284,342	2,279,511	2,279,511	0.21	0.00
	18-25	1,177	2,942,126	2,943,283	2,943,283	-0.04	0.00
	26-34	343	3,190,481	3,263,612	3,263,612	-2.24	0.00
	35-49	517	5,196,974	5,124,549	5,124,549	1.41	0.00
	50-64	198	4,196,966	4,486,014	4,486,014	-6.44	0.00
	65+	146	3,041,955	2,755,876	2,755,876	10.38	0.00
Race	**White**	2,649	15,601,222	16,945,306	16,945,306	-7.93	0.00
	Black or African American	465	2,628,668	2,483,873	2,483,873	5.83	0.00
	Other	511	2,622,953	1,423,665	1,423,665	84.24	0.00
Hispanicity	**Hispanic or Latino**	1,660	7,609,601	7,481,864	7,481,864	1.71	0.00
	Non-Hispanic or Latino	1,965	13,243,243	13,370,980	13,370,980	-0.96	0.00
Gender	**Male**	1,777	10,164,233	10,157,252	10,157,252	0.07	0.00
	Female	1,848	10,688,611	10,695,592	10,695,592	-0.07	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.46 2012 NSDUH Slippage Rates: UTAH

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		926	2,214,352	2,214,352	2,214,352	0.00	0.00
Quarter	**Quarter 1**	206	549,792	549,792	549,792	0.00	0.00
	Quarter 2	239	552,364	552,365	552,365	-0.00	0.00
	Quarter 3	226	554,900	554,900	554,900	0.00	0.00
	Quarter 4	255	557,295	557,295	557,295	0.00	0.00
Age Group	**12-17**	286	271,013	272,004	272,004	-0.36	0.00
	18-25	307	361,279	363,798	363,798	-0.69	0.00
	26-34	123	396,299	396,762	396,762	-0.12	0.00
	35-49	106	504,316	500,343	500,343	0.79	0.00
	50-64	67	439,409	414,570	414,570	5.99	-0.00
	65+	37	242,036	266,874	266,874	-9.31	0.00
Race	**White**	852	2,036,071	2,045,611	2,045,611	-0.47	0.00
	Black or African American	9	20,853	25,879	25,879	-19.42	0.00
	Other	65	157,428	142,862	142,862	10.20	0.00
Hispanicity	**Hispanic or Latino**	185	277,215	269,478	269,478	2.87	0.00
	Non-Hispanic or Latino	741	1,937,136	1,944,874	1,944,874	-0.40	-0.00
Gender	**Male**	444	1,095,875	1,101,009	1,101,009	-0.47	0.00
	Female	482	1,118,476	1,113,342	1,113,342	0.46	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.47 2012 NSDUH Slippage Rates: VERMONT

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		885	541,583	541,583	541,583	-0.00	0.00
Quarter	**Quarter 1**	202	135,315	135,315	135,315	0.00	0.00
	Quarter 2	261	135,379	135,379	135,379	-0.00	0.00
	Quarter 3	259	135,433	135,433	135,433	-0.00	0.00
	Quarter 4	163	135,456	135,456	135,456	0.00	0.00
Age Group	**12-17**	282	45,027	45,420	45,420	-0.87	0.00
	18-25	320	75,154	73,055	73,055	2.87	0.00
	26-34	63	61,423	63,130	63,130	-2.70	0.00
	35-49	110	121,539	119,148	119,148	2.01	0.00
	50-64	72	157,946	145,719	145,719	8.39	0.00
	65+	38	80,494	95,111	95,111	-15.37	0.00
Race	**White**	837	513,765	519,131	519,131	-1.03	0.00
	Black or African American	8	5,496	5,267	5,267	4.36	0.00
	Other	40	22,322	17,185	17,185	29.89	0.00
Hispanicity	**Hispanic or Latino**	22	7,282	8,148	8,148	-10.63	0.00
	Non-Hispanic or Latino	863	534,300	533,434	533,434	0.16	0.00
Gender	**Male**	439	265,224	265,224	265,224	-0.00	0.00
	Female	446	276,359	276,359	276,359	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.48 2012 NSDUH Slippage Rates: VIRGINIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		894	6,735,698	6,735,698	6,735,698	0.00	0.00
Quarter	**Quarter 1**	220	1,676,594	1,676,594	1,676,594	0.00	0.00
	Quarter 2	252	1,681,743	1,681,743	1,681,743	0.00	0.00
	Quarter 3	205	1,686,584	1,686,584	1,686,584	0.00	0.00
	Quarter 4	217	1,690,778	1,690,778	1,690,778	0.00	0.00
Age Group	**12-17**	323	623,960	619,042	619,042	0.79	0.00
	18-25	264	869,496	891,542	891,542	-2.47	0.00
	26-34	99	972,615	966,143	966,143	0.67	0.00
	35-49	102	1,599,104	1,626,937	1,626,937	-1.71	0.00
	50-64	66	1,727,822	1,597,459	1,597,459	8.16	-0.00
	65+	40	942,702	1,034,575	1,034,575	-8.88	0.00
Race	**White**	633	4,700,130	4,883,476	4,883,476	-3.75	0.00
	Black or African American	144	1,333,500	1,272,186	1,272,186	4.82	0.00
	Other	117	702,068	580,036	580,036	21.04	0.00
Hispanicity	**Hispanic or Latino**	77	535,154	513,221	513,221	4.27	0.00
	Non-Hispanic or Latino	817	6,200,544	6,222,477	6,222,477	-0.35	-0.00
Gender	**Male**	422	3,228,718	3,228,718	3,228,718	0.00	0.00
	Female	472	3,506,980	3,506,980	3,506,980	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.49 2012 NSDUH Slippage Rates: WASHINGTON

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		928	5,736,136	5,736,136	5,736,136	-0.00	0.00
Quarter	**Quarter 1**	268	1,427,477	1,427,477	1,427,477	-0.00	0.00
	Quarter 2	169	1,432,111	1,432,111	1,432,111	0.00	-0.00
	Quarter 3	263	1,436,429	1,436,429	1,436,430	-0.00	-0.00
	Quarter 4	228	1,440,118	1,440,118	1,440,118	-0.00	0.00
Age Group	**12-17**	300	527,766	528,812	528,812	-0.20	-0.00
	18-25	307	730,498	737,911	737,911	-1.00	0.00
	26-34	87	836,406	850,975	850,975	-1.71	0.00
	35-49	137	1,377,341	1,354,312	1,354,312	1.70	-0.00
	50-64	66	1,543,331	1,375,368	1,375,368	12.21	-0.00
	65+	31	720,795	888,758	888,758	-18.90	0.00
Race	**White**	724	4,590,573	4,747,515	4,747,515	-3.31	0.00
	Black or African American	33	205,207	206,654	206,654	-0.70	-0.00
	Other	171	940,356	781,967	781,967	20.26	0.00
Hispanicity	**Hispanic or Latino**	126	578,060	575,723	575,723	0.41	0.00
	Non-Hispanic or Latino	802	5,158,076	5,160,413	5,160,413	-0.05	0.00
Gender	**Male**	476	2,822,981	2,822,981	2,822,981	-0.00	-0.00
	Female	452	2,913,155	2,913,155	2,913,155	-0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.50 2012 NSDUH Slippage Rates: WEST VIRGINIA

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		976	1,574,171	1,574,171	1,574,171	0.00	0.00
Quarter	**Quarter 1**	194	393,313	393,313	393,313	0.00	0.00
	Quarter 2	315	393,418	393,418	393,418	0.00	-0.00
	Quarter 3	188	393,617	393,617	393,617	0.00	0.00
	Quarter 4	279	393,823	393,823	393,823	0.00	0.00
Age Group	**12-17**	313	130,371	131,131	131,131	-0.58	-0.00
	18-25	354	195,294	189,192	189,192	3.23	0.00
	26-34	66	188,884	192,823	192,823	-2.04	0.00
	35-49	113	345,810	350,672	350,672	-1.39	0.00
	50-64	73	410,829	406,985	406,985	0.94	0.00
	65+	57	302,984	303,368	303,368	-0.13	0.00
Race	**White**	890	1,478,055	1,490,548	1,490,549	-0.84	-0.00
	Black or African American	41	49,290	50,778	50,778	-2.93	0.00
	Other	45	46,826	32,845	32,845	42.57	0.00
Hispanicity	**Hispanic or Latino**	20	19,773	17,974	17,974	10.01	0.00
	Non-Hispanic or Latino	956	1,554,398	1,556,197	1,556,197	-0.12	0.00
Gender	**Male**	495	767,075	767,075	767,075	0.00	0.00
	Female	481	807,096	807,096	807,096	0.00	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.51 2012 NSDUH Slippage Rates: WISCONSIN

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		875	4,786,617	4,786,617	4,786,617	0.00	0.00
Quarter	**Quarter 1**	223	1,194,235	1,194,235	1,194,235	0.00	0.00
	Quarter 2	241	1,195,733	1,195,733	1,195,733	0.00	0.00
	Quarter 3	256	1,197,476	1,197,476	1,197,476	0.00	0.00
	Quarter 4	155	1,199,173	1,199,173	1,199,173	0.00	0.00
Age Group	**12-17**	273	448,603	450,470	450,470	-0.41	0.00
	18-25	322	610,171	615,758	615,758	-0.91	0.00
	26-34	88	647,906	649,743	649,743	-0.28	0.00
	35-49	106	1,093,187	1,092,759	1,092,759	0.04	0.00
	50-64	48	1,097,475	1,181,693	1,181,693	-7.13	0.00
	65+	38	889,275	796,194	796,194	11.69	0.00
Race	**White**	747	4,225,631	4,293,447	4,293,447	-1.58	0.00
	Black or African American	61	285,513	276,619	276,619	3.22	0.00
	Other	67	275,473	216,550	216,550	27.21	0.00
Hispanicity	**Hispanic or Latino**	85	228,221	251,411	251,411	-9.22	0.00
	Non-Hispanic or Latino	790	4,558,396	4,535,205	4,535,205	0.51	0.00
Gender	**Male**	443	2,351,239	2,352,263	2,352,263	-0.04	0.00
	Female	432	2,435,378	2,434,353	2,434,353	0.04	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table H.52 2012 NSDUH Slippage Rates: WYOMING

Domain		*n*	Initial Total (I)[1]	Final Total (F)[2]	Census Total (C)	(I-C)/C%	(F-C)/C%
Total		928	474,248	474,248	474,248	0.00	0.00
Quarter	**Quarter 1**	229	117,747	117,747	117,747	0.00	0.00
	Quarter 2	203	118,296	118,296	118,296	0.00	0.00
	Quarter 3	257	118,844	118,844	118,844	0.00	0.00
	Quarter 4	239	119,361	119,361	119,361	0.00	0.00
Age Group	**12-17**	293	43,140	43,140	43,140	-0.00	0.00
	18-25	286	64,302	63,681	63,681	0.97	0.00
	26-34	98	69,515	70,690	70,690	-1.66	0.00
	35-49	141	105,231	102,096	102,096	3.07	0.00
	50-64	67	118,263	121,484	121,484	-2.65	0.00
	65+	43	73,797	73,158	73,158	0.87	0.00
Race	**White**	840	435,189	444,468	444,468	-2.09	0.00
	Black or African American	13	8,525	6,855	6,855	24.35	0.00
	Other	75	30,535	22,924	22,924	33.20	0.00
Hispanicity	**Hispanic or Latino**	104	42,156	40,588	40,588	3.86	0.00
	Non-Hispanic or Latino	824	432,092	433,660	433,660	-0.36	0.00
Gender	**Male**	434	239,359	239,994	239,994	-0.26	0.00
	Female	494	234,889	234,254	234,254	0.27	0.00

[1] WT1*...*WT13 (before person poststratification).
[2] WT1*...*WT14 (after person poststratification).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Appendix I: Evaluation of Calibration Weights: Weight Summary Statistics

Table I.1 2012 NSDUH Dwelling Unit–Level Weight Summary Statistics: United States, District of Columbia, and the 50 States

Domain	*n*	Before res.du.nr (WT1*...*WT7)[1]						After res.du.nr & Before res.du.ps (WT1*...*WT8)[1]						After res.du.ps (WT1*...*WT9)[1]					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
United States	153,873	30	366	517	829	3,566	1.44	40	401	602	955	6,317	1.45	9	417	644	1,043	9,546	1.54
Alaska	1,642	101	107	118	124	230	1.01	103	120	131	141	274	1.02	39	137	151	175	506	1.06
Alabama	2,141	286	634	652	846	886	1.02	551	708	748	940	1,100	1.02	213	761	891	1,036	2,893	1.07
Arkansas	2,090	186	389	511	541	777	1.03	210	464	555	601	871	1.04	76	489	552	639	2,351	1.07
Arizona	1,928	689	723	763	799	2,813	1.20	721	787	859	987	3,189	1.21	192	945	1,168	1,402	4,992	1.20
California	6,852	610	1,267	1,333	1,442	2,999	1.01	956	1,498	1,638	1,780	3,494	1.02	455	1,608	1,817	2,035	8,444	1.10
Colorado	2,201	546	567	641	795	3,062	1.32	548	639	724	939	2,467	1.19	143	730	861	1,032	3,235	1.12
Connecticut	2,107	383	444	451	579	642	1.02	471	507	535	695	2,128	1.07	113	535	595	808	5,276	1.16
District of Columbia	3,327	37	52	54	57	183	1.05	43	59	67	70	176	1.07	9	68	79	94	281	1.09
Delaware	2,008	72	126	130	143	250	1.01	81	142	149	167	231	1.01	33	143	168	194	891	1.13
Florida	8,516	181	547	636	650	2,119	1.05	371	691	734	781	1,528	1.04	319	757	859	961	3,567	1.06
Georgia	1,796	1,372	1,421	1,476	1,701	2,919	1.01	1,405	1,604	1,742	1,929	4,030	1.02	310	1,635	1,912	2,218	9,424	1.09
Hawaii	2,239	120	128	136	155	232	1.05	133	152	170	210	385	1.07	49	157	181	230	1,330	1.14
Iowa	2,022	456	476	675	704	1,674	1.05	466	521	704	780	1,824	1.06	184	523	581	704	2,421	1.08
Idaho	1,821	30	236	246	294	474	1.03	120	251	275	312	497	1.03	97	281	312	342	1,653	1.07
Illinois	7,678	50	405	418	455	630	1.01	192	495	545	607	1,874	1.05	269	532	604	682	3,166	1.05
Indiana	1,921	287	955	1,131	1,229	1,347	1.01	486	1,071	1,237	1,328	5,490	1.03	343	1,161	1,317	1,451	6,654	1.07
Kansas	1,977	93	443	460	472	602	1.02	136	485	511	560	953	1.02	179	500	545	597	2,194	1.07
Kentucky	2,202	633	656	678	758	1,006	1.03	633	704	777	867	1,296	1.03	228	698	775	880	2,071	1.04
Louisiana	1,977	562	607	656	839	899	1.03	600	669	713	904	1,104	1.03	239	768	890	1,016	3,677	1.05
Massachusetts	2,208	760	788	881	892	1,456	1.01	813	925	1,059	1,106	2,213	1.02	180	1,028	1,155	1,247	5,531	1.11
Maryland	1,802	309	855	892	921	981	1.00	890	1,078	1,108	1,190	1,613	1.01	280	1,085	1,180	1,352	4,429	1.06
Maine	2,585	166	172	176	202	237	1.01	173	187	195	210	470	1.02	37	197	217	235	923	1.06
Michigan	7,826	40	312	382	418	1,017	1.03	40	383	459	512	805	1.03	42	425	489	557	2,192	1.04
Minnesota	1,975	782	806	845	1,098	2,389	1.10	817	877	988	1,198	2,673	1.11	360	881	1,054	1,246	4,679	1.11
Missouri	2,188	855	888	906	1,042	1,237	1.01	908	970	1,038	1,142	1,377	1.01	268	928	1,052	1,167	5,701	1.06

(continued)

Table I.1 2012 NSDUH Dwelling Unit–Level Weight Summary Statistics: United States, District of Columbia, and the 50 States (continued)

Domain	*n*	Before res.du.nr (WT1*...*WT7)[1]						After res.du.nr & Before res.du.ps (WT1*...*WT8)[1]						After res.du.ps (WT1*...*WT9)[1]					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Mississippi	1,951	128	459	506	518	665	1.01	255	501	536	583	764	1.01	97	518	595	678	2,006	1.07
Montana	2,415	127	129	137	139	332	1.02	130	140	144	150	225	1.01	49	156	167	184	582	1.03
North Carolina	1,990	1,333	1,495	1,631	1,681	3,566	1.01	1,336	1,626	1,748	1,896	3,816	1.02	346	1,509	1,799	2,151	9,546	1.14
North Dakota	2,461	77	80	86	100	142	1.01	78	85	92	103	121	1.01	33	100	114	128	417	1.08
Nebraska	2,018	272	282	293	317	333	1.00	277	304	317	337	410	1.01	91	322	344	380	1,927	1.08
New Hampshire	2,191	178	184	190	198	390	1.01	180	209	217	227	413	1.01	38	217	232	255	1,118	1.06
New Jersey	1,935	1,042	1,117	1,148	1,498	1,857	1.03	1,089	1,264	1,398	1,699	6,317	1.06	570	1,343	1,635	1,961	7,096	1.11
New Mexico	1,889	74	255	275	286	424	1.03	245	286	296	329	479	1.02	65	350	398	459	1,644	1.09
Nevada	1,721	153	257	354	391	469	1.04	153	344	394	429	837	1.04	155	438	543	669	2,466	1.18
New York	9,115	48	494	556	581	1,629	1.02	252	636	720	868	2,305	1.08	259	641	732	893	7,959	1.22
Ohio	9,023	34	403	415	431	525	1.01	193	450	464	494	879	1.01	166	471	497	546	1,777	1.03
Oklahoma	2,173	178	504	560	593	884	1.02	374	544	617	704	1,639	1.03	174	580	689	805	2,585	1.09
Oregon	2,019	583	611	639	695	778	1.01	618	655	723	799	1,047	1.02	211	690	757	836	2,877	1.07
Pennsylvania	8,453	135	395	440	523	799	1.02	149	480	555	606	1,790	1.06	240	483	580	666	4,776	1.09
Rhode Island	1,957	127	137	161	193	200	1.02	143	161	186	213	448	1.03	41	177	206	242	1,027	1.15
South Carolina	2,374	539	576	624	647	759	1.01	560	637	683	722	1,185	1.01	185	683	752	837	1,942	1.05
South Dakota	2,031	93	114	134	159	170	1.02	95	116	147	171	281	1.04	40	142	163	180	537	1.06
Tennessee	1,929	581	851	1,157	1,194	2,421	1.02	835	963	1,235	1,316	2,381	1.03	246	1,155	1,294	1,415	3,727	1.06
Texas	6,792	467	1,061	1,099	1,133	3,064	1.05	859	1,165	1,214	1,290	3,862	1.06	343	1,185	1,312	1,455	6,541	1.09
Utah	1,474	403	426	444	497	574	1.01	417	451	481	535	617	1.01	85	536	598	672	1,831	1.07
Virginia	2,027	561	1,271	1,329	1,370	1,432	1.00	993	1,427	1,504	1,552	2,338	1.01	326	1,354	1,485	1,646	4,387	1.05
Vermont	2,317	66	79	87	89	134	1.04	70	91	98	106	320	1.05	18	91	103	128	553	1.13
Washington	2,078	595	957	1,009	1,175	1,404	1.02	659	1,066	1,162	1,365	1,644	1.03	640	1,112	1,256	1,428	5,924	1.08
Wisconsin	1,890	805	837	935	1,079	1,175	1.02	816	882	1,031	1,152	1,631	1.03	293	1,046	1,146	1,308	3,693	1.06
West Virginia	2,399	84	223	295	318	372	1.03	156	249	325	379	504	1.04	35	259	319	391	1,129	1.09
Wyoming	2,222	69	72	88	92	116	1.01	73	84	95	99	186	1.01	20	94	104	115	347	1.05

[1] WT1*…*WT7 are design-based weight components; nr = nonresponse adjustment; ps = poststratification adjustment.

[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.

[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table I.2 2012 NSDUH Selected Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States

Domain	*n*	Before sel.per.ps (WT1*...*WT11)[1]						After sel.per.ps (WT1*...*WT12)[1]					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
United States	87,656	19	677	1,334	3,483	75,644	2.89	4	650	1,330	3,466	83,336	2.94
Alaska	1,076	66	176	259	764	3,894	2.02	28	181	281	711	3,455	2.09
Alabama	1,145	236	1,076	1,474	5,189	19,613	1.93	282	1,053	1,544	5,234	20,719	2.10
Arkansas	1,212	83	569	793	3,037	19,370	2.14	18	539	842	3,098	21,749	2.38
Arizona	1,139	302	1,307	2,111	6,364	48,558	2.38	172	1,284	2,147	5,725	38,233	2.54
California	4,779	519	2,108	3,006	9,534	75,644	2.07	437	2,109	3,134	9,303	82,766	2.09
Colorado	1,188	154	1,018	1,508	5,573	34,687	2.23	84	954	1,540	5,529	52,100	2.31
Connecticut	1,261	114	674	1,015	3,544	30,404	2.35	30	658	1,059	3,367	22,163	2.40
District of Columbia	1,125	19	101	239	812	3,518	2.32	4	103	215	728	3,470	2.38
Delaware	1,110	43	178	342	1,125	5,563	2.07	17	189	351	1,018	4,745	2.10
Florida	4,579	334	964	1,330	5,487	34,154	2.17	67	956	1,360	5,631	30,510	2.23
Georgia	1,144	444	2,305	3,337	10,689	59,251	1.96	312	2,323	3,312	10,810	41,529	1.88
Hawaii	1,285	69	243	445	1,335	8,573	2.08	39	233	447	1,181	6,563	2.15
Iowa	1,137	223	629	942	2,998	21,037	2.22	98	597	906	2,938	19,002	2.49
Idaho	1,136	98	343	470	1,479	4,938	1.87	35	335	506	1,555	7,455	2.20
Illinois	4,871	310	666	923	3,275	17,157	2.00	170	651	981	3,228	16,070	2.07
Indiana	1,171	380	1,457	2,076	7,400	68,551	2.14	107	1,478	2,095	7,073	33,793	1.97
Kansas	1,109	195	610	935	2,933	15,257	2.04	58	610	1,008	2,934	15,527	2.19
Kentucky	1,184	244	924	1,172	4,824	22,032	2.09	149	876	1,223	4,866	16,931	2.08
Louisiana	1,100	357	1,138	1,544	4,948	17,111	1.96	154	1,056	1,632	4,886	23,499	2.10
Massachusetts	1,253	210	1,239	1,874	6,738	38,702	2.17	85	1,293	2,004	6,588	47,604	2.28
Maryland	1,074	305	1,323	1,796	6,838	31,262	2.00	284	1,248	1,938	6,760	37,356	2.25
Maine	1,134	39	261	334	1,397	7,926	2.39	8	254	354	1,520	6,742	2.41
Michigan	4,606	55	548	749	2,667	10,908	2.06	47	549	791	2,615	26,027	2.15
Minnesota	1,092	379	1,136	1,718	5,017	32,719	2.33	96	1,082	1,788	4,588	29,298	2.67
Missouri	1,149	398	1,279	1,832	6,435	40,189	2.01	114	1,292	1,935	6,582	31,554	2.08

(continued)

Table I.2 2012 NSDUH Selected Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States (continued)

		Before sel.per.ps (WT1*...*WT11)[1]						After sel.per.ps (WT1*...*WT12)[1]					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
Mississippi	1,100	103	687	995	3,098	12,749	2.03	21	620	999	3,113	14,651	2.26
Montana	1,109	57	196	286	1,064	3,846	2.11	49	193	298	971	5,628	2.66
North Carolina	1,117	429	1,958	3,006	10,385	56,513	2.25	132	1,802	3,055	9,969	83,336	2.41
North Dakota	1,156	40	148	282	686	2,858	1.96	27	134	286	698	3,783	1.98
Nebraska	1,170	169	391	524	2,041	11,810	2.07	37	387	575	1,874	11,061	2.25
New Hampshire	1,259	62	257	325	1,382	4,709	2.13	23	256	342	1,437	6,487	2.20
New Jersey	1,155	576	1,740	2,803	9,005	47,406	2.18	450	1,668	2,817	9,376	56,420	2.37
New Mexico	1,101	68	474	662	2,066	10,620	2.18	64	477	693	2,066	12,003	2.38
Nevada	1,134	168	528	898	2,771	30,652	2.34	123	518	911	2,563	16,963	2.37
New York	5,267	308	897	1,315	4,467	47,814	2.30	234	856	1,363	4,385	38,090	2.35
Ohio	4,827	192	601	789	3,013	15,894	2.01	81	599	826	3,092	12,273	2.02
Oklahoma	1,189	204	779	1,162	4,150	17,967	2.01	134	739	1,150	3,992	14,076	2.09
Oregon	1,165	290	828	1,154	4,578	23,037	2.11	116	803	1,222	3,972	24,548	2.25
Pennsylvania	4,705	247	635	989	3,502	19,968	2.11	124	622	1,018	3,436	15,262	2.10
Rhode Island	1,131	53	247	356	1,210	7,822	2.12	13	228	372	1,116	9,729	2.26
South Carolina	1,171	226	989	1,477	4,972	21,268	1.97	77	952	1,576	4,352	18,966	2.17
South Dakota	1,113	50	191	270	890	4,623	2.03	25	188	272	996	3,249	1.89
Tennessee	1,105	264	1,410	1,882	7,758	44,286	2.14	61	1,387	2,069	7,425	28,573	2.19
Texas	4,612	416	1,476	1,980	6,458	64,077	2.05	394	1,462	2,049	6,641	38,094	1.96
Utah	1,099	105	792	1,079	2,817	12,752	1.84	29	755	1,119	3,151	11,370	1.81
Virginia	1,095	372	1,712	2,801	8,687	43,656	2.04	128	1,707	2,799	9,297	37,275	2.07
Vermont	1,136	19	131	180	693	4,525	2.38	9	123	193	681	3,359	2.31
Washington	1,218	746	1,369	1,924	7,051	61,037	2.31	350	1,293	2,018	6,617	44,736	2.29
Wisconsin	1,098	298	1,186	1,665	6,336	31,882	2.36	344	1,171	1,660	5,839	32,089	2.44
West Virginia	1,217	62	341	472	1,856	7,129	2.36	35	350	486	2,250	9,236	2.21
Wyoming	1,148	29	131	183	602	3,455	2.01	8	133	200	542	2,475	2.07

[1] WT1*...*WT11 and WT1*...*WT12 used demographic variables from screener data; ps = poststratification adjustment.

[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.

[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table I.3 2012 NSDUH Respondent Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States

		Before res.per.nr (WT1*...*WT12)[1]						After res.per.nr (WT1*...*WT13)[1]						Before res.per.ps (WT1*...*WT13)[2]						Final Weight After res.per.ps (WT1*...*WT14)[2]					
Domain	***n***	**Min**	**Q1[3]**	**Med**	**Q3[3]**	**Max**	**UWE[4]**	**Min**	**Q1[3]**	**Med**	**Q3[3]**	**Max**	**UWE[4]**	**Min**	**Q1[3]**	**Med**	**Q3[3]**	**Max**	**UWE[4]**	**Min**	**Q1[3]**	**Med**	**Q3[3]**	**Max**	**UWE[4]**
United States	68,309	4	623	1,263	3,147	66,615	3.00	4	757	1,556	4,107	112,786	3.44	4	757	1,556	4,107	112,786	3.44	1	726	1,549	4,069	133,926	3.55
Alaska	829	28	181	262	688	3,455	2.18	33	219	346	892	7,281	2.30	33	219	346	892	7,281	2.30	13	220	357	888	5,841	2.33
Alabama	901	333	1,037	1,479	4,909	20,719	2.14	333	1,275	1,865	6,277	31,521	2.34	333	1,275	1,865	6,277	31,521	2.34	71	1,295	1,873	6,084	38,506	2.52
Arkansas	913	39	523	792	2,683	17,515	2.33	39	647	1,038	3,669	21,376	2.59	39	647	1,038	3,669	21,376	2.59	19	646	1,072	3,514	30,876	2.74
Arizona	922	172	1,260	2,039	5,458	38,233	2.51	176	1,466	2,584	6,796	55,524	2.79	176	1,466	2,584	6,796	55,524	2.79	62	1,427	2,491	6,624	69,374	2.96
California	3,608	500	2,053	2,934	8,747	66,615	2.12	657	2,545	3,819	11,883	112,786	2.50	657	2,545	3,819	11,883	112,786	2.50	251	2,545	4,018	11,437	133,926	2.61
Colorado	927	84	944	1,488	5,273	52,100	2.37	88	1,147	1,887	6,714	55,395	2.47	88	1,147	1,887	6,714	55,395	2.47	30	1,166	1,859	6,449	44,336	2.57
Connecticut	964	30	642	1,024	3,082	22,163	2.45	30	825	1,275	4,230	38,798	2.74	30	825	1,275	4,230	38,798	2.74	7	783	1,344	4,040	28,699	2.74
District of Columbia	962	4	97	201	654	3,470	2.46	4	110	228	797	4,426	2.59	4	110	228	797	4,426	2.59	1	108	231	793	5,289	2.71
Delaware	893	17	190	355	986	4,745	2.10	17	238	426	1,234	6,524	2.20	17	238	426	1,234	6,524	2.20	12	234	433	1,242	8,502	2.30
Florida	3,544	67	940	1,295	5,130	30,510	2.31	87	1,092	1,623	6,586	48,148	2.74	87	1,092	1,623	6,586	48,148	2.74	46	1,095	1,675	6,257	44,131	2.83
Georgia	885	312	2,245	3,155	10,346	41,529	1.91	312	2,712	3,962	13,030	91,132	2.27	312	2,712	3,962	13,030	91,132	2.27	143	2,713	4,029	13,143	97,222	2.26
Hawaii	938	49	226	415	1,081	6,563	2.27	60	284	542	1,545	9,210	2.46	60	284	542	1,545	9,210	2.46	25	278	556	1,503	10,176	2.55
Iowa	900	98	581	877	2,756	18,917	2.56	126	678	1,113	3,744	26,336	2.73	126	678	1,113	3,744	26,336	2.73	56	677	1,108	3,762	52,057	3.09
Idaho	921	35	330	488	1,485	7,455	2.25	36	373	596	1,828	11,138	2.46	36	373	596	1,828	11,138	2.46	24	382	610	1,824	10,969	2.48
Illinois	3,672	170	638	948	3,035	14,206	2.11	170	793	1,236	4,115	31,833	2.41	170	793	1,236	4,115	31,833	2.41	73	809	1,258	4,100	43,209	2.45
Indiana	911	107	1,441	1,997	6,383	33,793	2.02	114	1,762	2,517	8,149	41,439	2.26	114	1,762	2,517	8,149	41,439	2.26	37	1,725	2,483	8,299	51,485	2.40
Kansas	912	58	592	964	2,669	15,527	2.24	58	664	1,138	3,259	22,705	2.53	58	664	1,138	3,259	22,705	2.53	36	656	1,142	3,148	47,503	2.87
Kentucky	927	189	867	1,203	4,524	16,931	2.14	197	1,046	1,526	5,829	25,543	2.41	197	1,046	1,526	5,829	25,543	2.41	59	1,055	1,574	5,308	33,713	2.57
Louisiana	901	154	1,037	1,546	4,345	21,330	2.13	219	1,174	1,799	5,451	29,154	2.36	219	1,174	1,799	5,451	29,154	2.36	47	1,212	1,840	5,309	29,752	2.35
Massachusetts	955	85	1,243	1,839	6,356	38,627	2.29	131	1,499	2,477	7,793	65,152	2.55	131	1,499	2,477	7,793	65,152	2.55	122	1,583	2,578	7,946	46,935	2.46
Maryland	874	284	1,228	1,807	6,125	37,356	2.34	285	1,456	2,185	7,320	57,505	2.68	285	1,456	2,185	7,320	57,505	2.68	145	1,421	2,301	7,165	52,380	2.75
Maine	938	8	252	351	1,442	5,093	2.45	8	295	414	1,803	8,390	2.58	8	295	414	1,803	8,390	2.58	4	288	419	1,825	19,351	2.82
Michigan	3,655	47	542	778	2,504	26,027	2.18	49	653	952	3,162	31,584	2.36	49	653	952	3,162	31,584	2.36	46	649	968	3,236	27,645	2.32
Minnesota	902	96	1,069	1,712	4,395	29,298	2.75	96	1,185	2,028	5,812	63,117	2.80	96	1,185	2,028	5,812	63,117	2.80	21	1,177	2,057	5,752	55,395	2.91
Missouri	915	114	1,260	1,818	5,759	31,554	2.15	188	1,424	2,129	7,161	50,247	2.54	188	1,424	2,129	7,161	50,247	2.54	38	1,448	2,212	7,104	45,627	2.59

(continued)

Table I.3 2012 NSDUH Respondent Person-Level Weight Summary Statistics: United States, District of Columbia, and the 50 States (continued)

Domain	n	Before res.per.nr (WT1*...*WT12)[1]						After res.per.nr (WT1*...*WT13)[1]						Before res.per.ps (WT1*...*WT13)[2]						Final Weight After res.per.ps (WT1*...*WT14)[2]					
		Min	Q1[3]	Med	Q3[3]	Max	UWE[4]	Min	Q1[3]	Med	Q3[3]	Max	UWE[4]	Min	Q1[3]	Med	Q3[3]	Max	UWE[4]	Min	Q1[3]	Med	Q3[3]	Max	UWE[4]
Mississippi	901	21	616	971	2,859	14,651	2.29	21	724	1,117	3,583	20,017	2.60	21	724	1,117	3,583	20,017	2.60	5	737	1,184	3,361	31,890	2.76
Montana	876	68	189	287	947	5,628	2.75	72	234	364	1,289	9,217	2.65	72	234	364	1,289	9,217	2.65	21	239	370	1,301	7,089	2.65
North Carolina	917	132	1,774	2,845	8,865	49,012	2.45	210	2,134	3,342	11,433	82,293	2.72	210	2,134	3,342	11,433	82,293	2.72	170	2,056	3,355	11,182	86,606	2.84
North Dakota	895	27	126	258	647	3,783	2.05	33	151	319	869	5,345	2.32	33	151	319	869	5,345	2.32	7	150	311	857	6,490	2.38
Nebraska	940	37	380	543	1,683	11,061	2.37	39	427	635	2,227	14,370	2.68	39	427	635	2,227	14,370	2.68	8	419	652	2,289	25,155	2.72
New Hampshire	950	23	257	339	1,424	4,362	2.19	23	310	469	1,803	11,212	2.36	23	310	469	1,803	11,212	2.36	5	310	456	1,763	9,974	2.41
New Jersey	898	471	1,615	2,751	7,989	56,420	2.46	471	1,779	3,492	9,737	81,887	2.84	471	1,779	3,492	9,737	81,887	2.84	68	1,857	3,447	10,053	111,268	3.07
New Mexico	879	64	466	653	1,896	12,003	2.43	67	534	791	2,401	18,212	2.80	67	534	791	2,401	18,212	2.80	55	523	789	2,412	23,173	2.92
Nevada	903	123	500	867	2,377	16,963	2.45	126	597	1,069	2,846	26,443	2.73	126	597	1,069	2,846	26,443	2.73	30	553	1,073	2,792	32,722	3.17
New York	3,680	234	822	1,269	3,911	38,090	2.38	293	1,066	1,772	5,586	78,212	2.95	293	1,066	1,772	5,586	78,212	2.95	43	1,078	1,820	5,423	74,343	2.98
Ohio	3,687	88	590	802	3,000	10,538	2.05	128	731	1,053	3,944	17,838	2.24	128	731	1,053	3,944	17,838	2.24	77	745	1,079	3,962	16,099	2.26
Oklahoma	908	134	722	1,109	3,641	14,076	2.13	137	873	1,498	4,364	29,022	2.47	137	873	1,498	4,364	29,022	2.47	28	833	1,455	4,386	33,292	2.67
Oregon	923	116	779	1,205	3,904	24,548	2.29	116	917	1,480	4,816	42,723	2.42	116	917	1,480	4,816	42,723	2.42	72	919	1,565	4,861	35,124	2.43
Pennsylvania	3,580	124	606	954	3,084	14,714	2.19	141	758	1,224	4,315	24,788	2.43	141	758	1,224	4,315	24,788	2.43	23	771	1,220	4,256	29,514	2.46
Rhode Island	923	13	225	352	976	9,729	2.35	13	254	448	1,189	15,298	2.74	13	254	448	1,189	15,298	2.74	7	250	432	1,193	10,540	2.88
South Carolina	938	77	928	1,477	4,082	18,966	2.24	105	1,074	1,836	5,315	29,433	2.56	105	1,074	1,836	5,315	29,433	2.56	51	1,084	1,842	5,231	29,988	2.67
South Dakota	878	25	192	271	942	3,249	1.94	46	220	327	1,254	6,486	2.11	46	220	327	1,254	6,486	2.11	11	211	332	1,205	6,152	2.20
Tennessee	927	61	1,364	2,020	7,327	24,944	2.21	61	1,541	2,429	8,301	45,397	2.39	61	1,541	2,429	8,301	45,397	2.39	29	1,507	2,390	8,301	42,042	2.44
Texas	3,625	394	1,439	1,969	6,278	38,094	2.01	397	1,713	2,465	8,296	58,131	2.28	397	1,713	2,465	8,296	58,131	2.28	96	1,807	2,566	8,259	45,485	2.36
Utah	926	29	759	1,131	3,086	11,370	1.82	29	871	1,346	3,470	22,842	1.94	29	871	1,346	3,470	22,842	1.94	24	872	1,321	3,519	14,300	1.91
Virginia	894	128	1,650	2,651	8,253	37,275	2.15	142	1,952	3,028	10,622	84,610	2.44	142	1,952	3,028	10,622	84,610	2.44	43	1,886	3,021	10,440	75,208	2.46
Vermont	885	9	121	179	626	3,359	2.36	9	141	218	838	4,627	2.57	9	141	218	838	4,627	2.57	3	146	223	817	4,473	2.60
Washington	928	350	1,274	1,899	6,268	35,511	2.30	353	1,589	2,484	8,395	52,570	2.57	353	1,589	2,484	8,395	52,570	2.57	172	1,593	2,576	8,451	51,231	2.54
Wisconsin	875	371	1,164	1,598	5,558	28,513	2.49	388	1,419	2,057	7,182	47,292	2.75	388	1,419	2,057	7,182	47,292	2.75	111	1,404	2,060	7,082	71,502	2.84
West Virginia	976	36	339	484	1,901	9,236	2.31	36	394	562	2,583	11,468	2.49	36	394	562	2,583	11,468	2.49	17	404	564	2,614	11,686	2.51
Wyoming	928	17	128	194	517	2,475	2.12	19	151	242	667	4,381	2.26	19	151	242	667	4,381	2.26	12	152	241	661	3,568	2.34

[1] WT1*...*WT12 and WT1*...*WT13 used demographic variables from screener data; nr = nonresponse adjustment.

[2] WT1*...*WT13 and WT1*...*WT14 used demographic variables from questionnaire data; ps = poststratification adjustment.

[3] Q1 and Q3 refer to the first and third quartile of the weight distribution.

[4] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

QUESTIONNAIRE DWELLING UNIT-LEVEL AND PERSON PAIR-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.107.005
Contract No. HHSS283201000003C
Deliverable No. 41

Authors:

Matthew Westlake
Patrick Chen
Harper Gordek

Project Director: Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2014

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

QUESTIONNAIRE DWELLING UNIT-LEVEL AND PERSON PAIR-LEVEL SAMPLING WEIGHT CALIBRATION

Prepared for the 2012 Methodological Resource Book

RTI Project No. 0212800.001.107.005
Contract No. HHSS283201000003C
Deliverable No. 41

Authors:

Matthew Westlake
Patrick Chen
Harper Gordek

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

February 2014

Preface

This report documents the method of weight calibration used for producing the final set of questionnaire dwelling unit (QDU) and pair weights for the National Survey on Drug Use and Health (NSDUH) data from 2012. The weighting team faced several challenges in this task and was able to address them by resorting to innovative modifications of certain basic statistical ideas, which are listed below.

- Under Brewer's method, high weights may occur due to small pair selection probabilities. In any calibration exercise, some treatment of extreme value (ev) in weights is needed, but there is a danger of introducing too much bias by over-treatment. In the generalized exponential model (GEM), which is described in detail in Chen et al. (2014), extreme value control is built in, but one needs to define suitable ev domains so that not too many evs are defined. If too many design variables are used to define ev domains, then each domain will be very sparse and will not be of much use in defining thresholds for ev. As in past surveys, a hierarchy of domains was defined using pair age (each pair member being in one of the three categories: 12 to 25, 26 to 49, and 50+) and number of persons aged 12 to 25 in the household, State, and clusters of States (see Section 5.2 for details).
- Control of extreme values in weights helps reduce instability of estimates to some extent, but there is a need for methods that do not introduce much bias. Following the famous suggestion of Hajek (1971) in his comments on Basu's fabled example of circus elephants, we performed ratio adjustment (a form of poststratification) to estimated totals obtained from the household data on the number of persons belonging to the pair domain of interest. This was implemented in a multivariate manner to get one set of final weights.
- In the absence of a suitable source of poststratification controls for the person pair-level weights and the household-level weights, the inherent two-phase nature of the survey design was capitalized upon to estimate these controls from the first phase of the large screener sample. The first-phase sample weight was poststratified to person-level U.S. Census Bureau counts to get more efficient estimated counts for pair and household data.
- The problem of multiplicities complicated the issue of providing one set of final weights. When dealing with person-level parameters involving drug-related behaviors among members of the same household, it is possible for an individual to manifest himself or herself in the pair sample through different pairs. To avoid overcounting, the pair weights have to be divided by multiplicity factors, which tend to be domain specific. For this reason, multiplicity factors for a key set of pair analysis domains also are produced along with a set of final calibrated pair weights.
- Missing items in the respondent questionnaire led to imputation for deriving pair relationships, multiplicity factors, and household counts for Hajek adjustments.

The calibration task described in this document has been in place, with minor modifications, since the 1999 version of NSDUH, which was then called the National Household Survey on Drug Abuse (NHSDA).[1] Results from this calibration applied to an earlier survey year were presented at the 2001 Joint Statistical Meetings. The procedures described in the proceedings papers from these presentations can serve as useful supplemental reference material on estimation in the presence of multiplicities and extreme weights (Chromy & Singh, 2001) and on GEM calibration of pair weights (Penne, Chen, & Singh, 2001). The experience of using GEM with person weights is described in an earlier proceedings paper (Chen, Penne, & Singh, 2000). This work was completed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. The authors would like to take this opportunity to thank a number of individuals for useful discussions and suggestions: Joe Gfroerer and Art Hughes of SAMHSA and Jim Chromy of RTI.

NSDUH Weighting Team
Ralph Folsom, Senior Advisor
Research Triangle Park, NC

[1] The National Household Survey on Drug Abuse (NHSDA) was renamed the National Survey on Drug Use and Health (NSDUH) in the 2002 survey year.

Table of Contents

Chapter **Page**

List of Terms and Abbreviations……...xi

1. Introduction 1

2. Questionnaire Dwelling Unit and Pair Selection Probabilities 7
   - 2.1 Pair Selection Probability 8
     - 2.1.1 Case I: DUs with $S \geq 2$ 8
     - 2.1.2 Case II: DUs with $S < 2$ 8
   - 2.2 Questionnaire Dwelling Unit Selection Probability 10

3. Brief Description of the Generalized Exponential Model 13

4. Predictor Variables for the Questionnaire Dwelling Unit and Pair Weight Calibration via the Generalized Exponential Model 15
   - 4.1 Questionnaire Dwelling Unit Weight Calibration 15
   - 4.2 Pair Weight Calibration 16

5. Definition of Extreme Weights 21
   - 5.1 Questionnaire Dwelling Unit Extreme Weight Definition 21
   - 5.2 Person Pair Extreme Weight Definition 22

6. Weight Calibration at Questionnaire Dwelling Unit and Pair Levels 25
   - 6.1 Phase I SDU-Level Weight Components 30
   - 6.2 QDU Weight Components 31
     - 6.2.1 QDU Weight Component #11: Inverse of Selection Probability of at Least One Person in the Dwelling Unit 31
     - 6.2.2 QDU Weight Component #12: Selected QDU Poststratification to SDU-Based Control Totals 31
     - 6.2.3 QDU Weight Component #13: Respondent QDU Nonresponse Adjustment 31
     - 6.2.4 QDU Weight Component #14: Respondent QDU Poststratification to SDU-Based Control Totals 32
     - 6.2.5 QDU Weight Component #15: Respondent QDU Extreme Value Adjustment 32
   - 6.3 Pair-Level Weight Components 32
     - 6.3.1 Pair Weight Component #11: Inverse of Selection Probability of a Person Pair in the Dwelling Unit 32
     - 6.3.2 Pair Weight Component #12: Selected Pair Poststratification to SDU-Based Control Totals 32
     - 6.3.3 Pair Weight Component #13: Respondent Pair Nonresponse Adjustment 32
     - 6.3.4 Pair Weight Component #14: Respondent Pair Poststratification to SDU-Based Control Totals 33
     - 6.3.5 Pair Weight Component #15: Respondent Pair Extreme Weight Adjustment 33

7. Evaluation of Calibration Weights 35
7.1 Response Rates 35
7.2 Proportions of Extreme Values and Outwinsors 35
7.3 Slippage Rates 36
7.4 Weight Adjustment Summary Statistics 36
7.5 Sensitivity Analysis of Drug Use Estimates 37
References 49

Appendix **Page**

A Technical Details about the Generalized Exponential Model A-1
B Derivation of Poststratification Control Totals B-1
C GEM Modeling Summary for the Questionnaire Dwelling Unit Weights C-1
C.1: Model Group 1: Northeast… C-11
C.2: Model Group 2: Midwest… C-19
C.3: Model Group 3: South… C-27
C.4: Model Group 4: West… C-35
D Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Response Rates D-1
E Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Proportions of Extreme Values and Outwinsors E-1
F Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Slippage Rates… F-1
G Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Weight Summary Statistics G-1
H GEM Modeling Summary for the Pair Weights H-1
H.1: Model Group 1: Northeast and South… H-5
H.2: Model Group 2: Midwest and West… H-13
I Evaluation of Calibration Weights: Pair-Level Response Rates I-1
J Evaluation of Calibration Weights: Pair-Level Proportions of Extreme Values and Outwinsors J-1
K Evaluation of Calibration Weights: Pair-Level Slippage Rates K-1
L Evaluation of Calibration Weights: Pair-Level Weight Summary Statistics L-1
M Pair Analysis Manual Excerpt M-1

List of Tables

Table		Page
1.1	2008–2012 NSDUH Sample Sizes	2
1.2	Pair Domains	3
2.1	Building Blocks of the QDU and Person Pair Samples: Dwelling Units and Persons in the 2008–2012 NSDUHs	11
6.1	Sample Size, by Model Group at QDU and Pair Levels	30
7.1	Estimates of Totals and SEs for Domains of Interest Based on QDU Sample: 2012	38
7.2a	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, by Mother Use: 2012	39
7.2b	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, by Father Use: 2012	40
7.3a	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Mother-Child (12 to 17) Pairs, by Mother Use: 2012	41
7.3b	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Father-Child (12 to 17) Pairs, by Father Use: 2012	42
7.4	Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	43
7.5	Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	44
7.6a	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Children about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	45
7.6b	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	46
7.7a	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	47
7.7b	Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012	48

C.1a 2012 QDU Weight GEM Modeling Summary (Model Group 1: Northeast) C-13
C.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast) C-14
C.2a 2012 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)..... C-21
C.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)..... C-22
C.3a 2012 QDU Weight GEM Modeling Summary (Model Group 3: South) C-29
C.3b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)..... C-30
C.4a 2012 QDU Weight GEM Modeling Summary (Model Group 4: West) C-37
C.4b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)..... C-38
D.1 2012 NSDUH QDU-Level Response Rates D-1
E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors... ..... E-1
E.2 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors... ..... E-3
F.1 2012 NSDUH QDU-Level Slippage Rates.....F-1
G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics G-1
G.2 2012 NSDUH Respondent QDU-Level Weight Summary Statistics... ..... G-3
H.1a 2012 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)..... H-7
H.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast and South) H-8
H.2a 2012 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)..... H-17
H.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest and West)..... H-18
I.1 2012 NSDUH Person Pair-Level Response Rates.....I-1
J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors... ..... J-1
J.2 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (res.pr.nr)... ..... J-3
J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (res.pr.ps and res.pr.ev) J-5
K.1 2012 NSDUH Respondent Pair-Level Slippage Rates K-1
L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics..... L-1
L.2 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr) L-4
L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev)..... L-6

List of Exhibits

Exhibit **Page**

1.1 QDU and Pair Sampling Weight Calibration Steps 4
4.1 Definitions of Levels for QDU-Level Calibration Modeling Variables 17
4.2 Definitions of Levels for Pair-Level Calibration Modeling Variables 18
6.1 Summary of 2012 NSDUH QDU Sample Weight Components 27
6.2 Summary of 2012 NSDUH Person Pair Sample Weight Components 28
6.3 U.S. Census Bureau Regions/Model Groups 29
C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables C-2
C.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights C-9
C.1.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 1: Northeast C-16
C.1.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 1: Northeast C-17
C.1.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 1: Northeast C-18
C.2.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 2: Midwest C-24
C.2.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 2: Midwest C-25
C.2.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest C-26
C.3.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 3: South C-32
C.3.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 3: South C-33
C.3.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 3: South C-34
C.4.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 4: West C-40
C.4.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 4: West C-41
C.4.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West C-42
H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables H-2
H.2 Covariates for 2012 NSDUH Pair Weights H-4
H.1.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 1: Northeast and South H-10
H.1.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 1: Northeast and South H-11
H.1.3 Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 1: Northeast and South H-12
H.1.4 Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 1: Northeast and South H-13

H.2.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West..... H-20
H.2.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 2: Midwest and West..... H-21
H.2.3 Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 2: Midwest and West..... H-22
H.2.4 Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 2: Midwest and West..... H-23

List of Terms and Abbreviations

DU	Dwelling unit.
Ev	Extreme value. See Sections 5.1 and 5.2 for more detail.
GEM	Generalized exponential model. See Chapter 3 for more detail.
Household-level person count	The number of pairs associated with a given domain in a given household. These counts are used as control totals in the poststratification step. See Chapter 11 in the editing and imputation report (Carpenter et al., 2014) for details on how these counts are created, and Chapter 4 for details on their use in poststratification.
IQR	Interquartile range.
Multiplicity factor	The number of pairs associated with a given respondent in a given domain. See Carpenter et al. (2014) for more detail.
Nr	Nonresponse.
Outwinsor	The proportion of weights trimmed after extreme value adjustment via winsorization.
Pair domain	A pair relationship where the target population is defined by one of the pair members, conditional on the attributes of the other pair member.
Pair relationship	The relationship between selected pair members.
Parent-child	A pair relationship where either both pair members identify the other as part of a parent-child relationship, or both pair members otherwise are determined to form a parent-child pair (either through other evidence or through imputation).
ps	Poststratification.
QDU	Questionnaire dwelling unit: a household where at least one member responded to the questionnaire.
res.pr.nr	Respondent pair nonresponse adjustment step. See Section 6.3.3 for more detail.
res.qdu.nr	Respondent questionnaire dwelling unit nonresponse adjustment step. See Section 6.2.3 for more detail.
res.pr.ev	Respondent pair extreme value adjustment step. See Section 6.3.5 for more detail.
res.qdu.ev	Respondent questionnaire dwelling unit extreme value adjustment step. See Section 6.2.5 for more detail.

res.pr.ps	Respondent pair poststratification adjustment step. See Section 6.3.4 for more detail.
res.qdu.ps	Respondent questionnaire dwelling unit poststratification adjustment step. See Section 6.2.4 for more detail.
SDU	Screener dwelling unit: a household where screener information is available.
sel.pr.ps	Selected person pair poststratification adjustment step. See Section 6.3.2 for more detail.
sel.qdu.ps	Selected questionnaire dwelling unit poststratification adjustment step. See Section 6.2.2 for more detail.
Sibling-sibling	A pair relationship where the pair members are siblings (either reported to be so, or otherwise determined to be so).
Spouse-spouse	A pair relationship where the pair members are either married or living together as though married (either reported to be so, or otherwise determined to be so).
SS	State sampling.
UWE	Unequal weighting effect. It refers to the contribution in the design effect due to unequal selection probability and is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights and n is the sample size.
Winsorization	A method of extreme value adjustment that replaces extreme values with the critical values used for defining low and high extreme values.

1. Introduction

Traditionally, most household surveys have been designed either to measure characteristics of the entire household or to focus on a randomly selected respondent from among those determined to be eligible for the survey. Selecting more than one person from the same household is generally avoided since persons from the same household often exhibit the same or similar characteristics and behavioral patterns. The intra-class correlation found among members of the same household leads to a clustering effect on the variance of estimates resulting in less precise estimates compared with estimates of the same sample size from a simple random sample. Selecting only one person per household avoids this clustering effect on the variance. The "one person per household" sampling approach, however, precludes the opportunity to gather information about the relationships among household members. In the National Survey on Drug Use and Health (NSDUH),[2] we allow for a richer analytic capability of a survey designed to ensure a positive pairwise probability of selection among all eligible household members in each sample household. Achieving positive probabilities for all pairs within sampled households permits unbiased estimation of the within-dwelling-unit component of variance. Besides providing efficient data collection, this sampling method also facilitates the study of the relationships of social behaviors among members of the same household. This report documents the methodology and development of calibrated weights for the second objective, the study of behavioral relationships among persons residing in the same household. The report also describes the development of questionnaire dwelling unit (QDU) weights, which are of independent interest for studying household-level characteristics and also are needed for producing household count estimates of the number of persons belonging to pair relationship domains for use as poststratification controls for pair weights.

NSDUH allows for estimating characteristics at the person level, pair level, and household or QDU level. This report describes the weight calibration methods used for the pair- and QDU-level respondents. As described in the person-level report, NSDUH is an annual survey of about 67,500 persons selected from the civilian, noninstitutionalized population aged 12 or older from all 50 States and the District of Columbia. Based on a composite size measure, States were geographically partitioned into roughly equal-sized regions according to population. The 42 smaller States and the District of Columbia were partitioned into 12 State sampling (SS) regions, whereas the eight largest States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions. Under a stratified design with States serving as the primary strata and SS regions serving as the secondary strata, census tracts, segments within census tracts, and dwelling units (DUs) within segments were each selected using probability proportional to size sampling. NSDUH is sometimes referred to as a two-phase sample where the first phase consisted of a large number of screener dwelling units (SDUs, about 200,000) selected to ensure that various age groups (five in all: 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50+) of eligible individuals were represented adequately in the second phase. Information collected from SDUs also provided estimates of population controls (as in two-phase sampling) for calibration at levels (such as pair and QDU) for which suitable U.S. Census Bureau-based controls were not available. The second phase consisted of

[2] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

the selection of zero, one, or two persons from each selected SDU using a modification of Brewer's method such that prescribed sampling rates for the five age groups in each State were achieved with high selection rates for youths (12 to 17) and young adults (18 to 25). Table 1.1 shows the eligible number of selected and responding SDUs, QDUs, pairs, and persons for each of the 5 years (2008–2012). The distribution of pair data for different pairs of age groups may vary considerably (see Chapter 2 for details). It is seen that for certain age group domains, the realized sample size may not be sufficient to yield reliable estimates. Also, there may be problems of extreme weights due to small pair selection probabilities under Brewer's method that may cause instability of estimates. These and some other estimation issues related to pair data are discussed below, along with some adopted solutions.

Table 1.1 2008–2012 NSDUH Sample Sizes

Sample Unit		2008	2009	2010	2011	2012
SDU	Selected	160,114	161,377	166,532	179,293	178,586
	Completed	142,159	142,933	147,010	156,048	153,873
QDU	Selected	58,942	58,288	58,702	61,441	60,621
	Completed	48,180	48,088	48,113	50,133	48,850
Pair	Selected	26,769	26,497	26,295	27,095	27,035
	Completed	19,748	19,919	19,691	19,976	19,459
Person	Selected	85,711	84,785	84,997	88,536	87,656
	Completed	67,928	68,007	67,804	70,109	68,309

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

First, we note that for studying drug-related behavioral relationships among members of the same household, pair data is required because the outcome variable generally is defined with respect to the specific other member selected from the household. However, the parameter of interest is generally at the person level and is not at the pair level. For example, in the parent-child pairs, one may be interested in the proportion of children that have used drugs in the past year who have parents that report talking to their child about drugs. Here the target population consists only of children, and not all possible pairs. Note that the pair-level (two persons per QDU) sample forms a subsample of the larger person-level (one or two persons per QDU) sample, with the QDUs themselves selected from the larger sample of SDUs. NSDUH has features of a two-phase design, which turns out to be useful for estimating calibration controls for poststratification of household-level weights and person pair-level weights. No other outside source is available for obtaining these controls. For this purpose, the screener-level household weights are poststratified to person-level census counts to obtain more efficient estimated controls for pair and household data.

In estimation for pair domains, two major problems arise: one is that of multiplicities because, for a given domain defined by the pair relationship, when the parameter of interest is at the person level, several pairs in the household could be associated with the same person, For example, analysts are interested in an outcome at the person level, the proportion of children who use drugs and whose parents report talking to them about drugs, where the focus is on the child in a parent-child pair. Several parent-child pairs in the household could be associated with the same child. If the household has two parents, the selected child has two inclusion possibilities

(one with each parent) in the set of all such parent-child pairs (Carpenter et al., 2014). The other problem is that of extreme weights that may arise due to small selection probabilities for certain pair age groups, which may lead to unstable estimates. Each of these issues is discussed in turn.

If several pairs in the household are associated with the same person, it is necessary to use the average measure of behavior relationships for each member, which gives rise to multiplicities. Thus, the pair weights need to be divided by the person-level multiplicity factors for each domain of interest, and, therefore, multiplicity factors need to be produced along with the final set of calibrated weights. Because it is not straightforward to create these multiplicities, analyses would have to be necessarily limited to pair relationships where the multiplicities were produced a priori. It was anticipated that analyses of interest would be limited to 14 pair domains, listed in Table 1.2. Since no multiplicity was necessary for the spouse-spouse/partner-partner pair relationships (by definition, each pair member could have only one partner or one spouse), multiplicity factors were produced for only 12 of these domains. Note that a single pair relationship might have two domains associated with it, since the parameter of interest might be associated with only one member of the pair (the "focus" member), and the multiplicity would differ depending upon which pair member was the focus member.

Table 1.2 Pair Domains

Pair Relationship	Focus
Parent-child: parent, child aged 12-14	Parent
Parent-child: parent, child aged 12-14	Child
Parent-child: parent, child aged 12-17	Parent
Parent-child: parent, child aged 12-17	Child
Parent-child: parent, child aged 12-20	Parent
Parent-child: parent, child aged 12-20	Child
Parent-child: parent, child aged 15-17	Parent
Parent-child: parent, child aged 15-17	Child
Sibling-sibling: older sibling 15-17, younger sibling 12-14	Older sibling
Sibling-sibling: older sibling 15-17, younger sibling 12-14	Younger sibling
Sibling-sibling: older sibling 18-25, younger sibling 12-17	Older sibling
Sibling-sibling: older sibling 18-25, younger sibling 12-17	Younger sibling
Spouse-spouse or partner-partner, with or without children	No multiplicity necessary
Spouse-spouse or partner-partner, with children aged 0-17	No multiplicity necessary

Some of the multiplicities, including counts of all possible pairs in a household for a given domain, were used for poststratification. Details are provided in Chapter 4. Prior editions of this report contained a chapter on editing and imputation of pair relationships, multiplicity factors, and household-level person counts for poststratification, but this information has been removed in the interest of consolidating the imputation documentation and can now be found in the editing and imputation report (Carpenter et al., 2014). Special consideration is required for analysis of pair-level data, and details can be found in *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* (Frechtel, Warren, & Porter, in press). See Appendix M for an excerpt from this manual on inferential population and multiplicities.

A resolution to the extreme weight problem is to use a Hajek-type modification (Hajek, 1971). This modification essentially entails calibration (like poststratification) to controls for the number of persons in households belonging to each domain of interest. These controls can be obtained from the larger sample of singles and pairs (i.e., one or two persons selected from DUs). Note, however, that the multiplicity factor, being domain specific, renders the calibration adjustment factor domain specific. This raises the question of finding one set of calibration weights for use with all domains or outcome variables. To get around this problem, we performed a multivariate calibration with respect to a key set of pair domains. This type of poststratification then was followed by a repeat poststratification to further control the extreme weights by imposing separate bound restrictions on the initially identified extreme weights.

The generalized exponential model (GEM) method (Folsom & Singh, 2000) was used for calibration of both QDU- and pair-level design weights through several steps of adjustment as shown in Exhibit 1.1. In GEM, treatment of extreme value (ev) weights is built in via the definition of lower and upper bounds for the extreme weights. For pair data, there was a problem defining suitable domains for defining extreme weights, as explained in the following paragraphs.

Exhibit 1.1 QDU and Pair Sampling Weight Calibration Steps

In dealing with extreme weights, it is assumed that they arise due to design (due to an imperfect frame, assignment of very small selection probabilities to some units, or a small weight adjustment factor after calibration) so that they make the sample representative of the population and, hence, do not introduce bias. The only problem is that they may lead to highly unstable

estimates similar to the problem of Basu's circus elephants[3] (Hajek, 1971). So, we need to perform some treatment (such as winsorization[4]) within suitably defined extreme weight domains such that these domains contain units possibly from different strata but with similar sample selection probabilities to avoid the occurrence of extreme weights due to a mix of different designs. The domains must be large enough (e.g., at least size 30) to be able to define extreme values according to the domain-specific weight distribution. Any extreme value treatment to increase precision of estimates would introduce some bias. However, this bias can be reduced considerably if the ev treatment is performed under calibration controls. This is what the built-in ev control in GEM tries to accomplish.

It follows that the definition of extreme weight domains should depend on factors that affect the selection probabilities of units in the sample, such as State- and age-specific sampling rates, segment selection probabilities, pair age-specific selection probabilities, and household composition. If one tries to define extreme weight domains by taking account of all these factors via cross-classification, it will lead to too many domains with insufficient observations. That is why it is difficult to define suitable extreme weight domains for pair data. In the case of person-level weights it was less difficult, since State by age group suitably captured the extreme weight domain requirements. The definition of extreme weight domains used in the 2012 survey was the same as the one used in the 1999–2011 surveys. The domains were defined as the cross-classification of State, pair age,[5] and number of persons aged 12 to 25 in a household. In particular, the pair age was defined by the age groups of each pair member according to the age categories of 12 to 25, 26 to 49, and 50 or older (resulting in six pair age categories), and the number of persons aged 12 to 25 were categorically defined as zero, one, and two or more. For more details, see Chapter 5.

[3] A circus owner had 50 elephants, and wanted to estimate the total weight to help him make arrangements for shipping. To save time, he only wanted to weigh Sambo (an average sized elephant), and use 50 times its weight as an estimate. However, the circus statistician, being highly conscious of the optimality and unbiasedness of the Horvitz-Thompson (HT) estimator, objected about the potential bias of his estimate because of the purposive selection. Instead, he suggested random selection of an elephant with a very high probability of 99/100 for Sambo, and the rest including Jumbo (the biggest in the herd) with probability 1/4900 each. The circus owner was very unhappy with the statistician's response of 100/99 times the Sambo's weight as the estimate if Sambo got selected in this random draw, and was outraged with the response of 4900 times the Jumbo's weight if Jumbo happened to get selected. It was obvious to the owner that this new estimator was extremely poor, although he didn't know anything about its unbiasedness. The story had an unhappy ending with the circus statistician losing his job. To alleviate the instability of the HT-estimator, Hajek suggested to multiply it by 50 divided by inverse of the selection probability, which reduces simply to 50 times the weight of the selected elephant.

[4] Winsorization is a method of extreme value adjustment that replaces extreme values with the critical values used for defining low and high extreme values.

[5] Pair age in this case should not be confused with the modeling term, which has a finer level breakdown.

2. Questionnaire Dwelling Unit and Pair Selection Probabilities

Similar to the 1999–2001 National Household Surveys on Drug Abuse (NHSDAs) and the 2002–2011 National Surveys on Drug Use and Health (NSDUHs),[6] the 2012 NSDUH had a two-phase design and used a computer-assisted interviewing (CAI) method. There were four stages of selection: census tracts, segments within census tracts, dwelling units (DUs) within segments, and persons within dwelling units. Any two survey eligible persons had some nonzero chance of being selected and, when both were selected, they formed a within household pair. This design feature is of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of an individual (in a family) relates to the drug use propensity of other members residing in the same dwelling unit (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

For the 1999–2001 surveys, the method used for selecting pairs was as follows. For a given DU, if the sum of the age-specific selection probabilities was larger than 2, then the individual person-selection probabilities were ratio adjusted downward to make the sum equal to 2. If the sum was less than 2, the difference between 2 and the sum of the probabilities was evenly distributed over three dummy persons so that the sum of the person probabilities was made to equal 2. Brewer's method was then applied to select a person pair. If the selected pair consisted of two real persons, then both persons were selected. If the selected pair consisted of one real person and one dummy person, then the real person was selected. If the selected pair consisted of two dummy persons, no one was selected from that DU.

Starting with the 2002 NSDUH and continuing through 2012, the pair-sampling algorithm was modified to increase the number of pairs selected in the sample. Dwelling units with the sum of person selection probabilities greater than or equal to 2 were treated the same as in previous survey years. However, DUs where the sum of person-level selection probabilities was less than 2 received a slightly different treatment that increased the chance for selecting a pair of real persons. Section 2.1 describes the selection process for both types of DUs.

Table 2.1 provides a summary of these NSDUH sampling units: eligible and completed screening dwelling units (SDUs), selected and completed questionnaire dwelling units (QDUs), selected and completed person interviews, and selected and completed person pairs, as well as their response rates. Using Brewer's method, zero, one, or two individuals were selected per household. Those SDUs where at least one person was selected were counted as the selected QDUs. A QDU where two persons were selected and both had completed interviews was considered to be a completed person pair. The table provides a breakdown by age group at the person level and age group by selection group (none, single, or pair) at the person pair level.

[6] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

2.1 Pair Selection Probability

2.1.1 Case I: DUs with $S \geq 2$

For a given DU, if the sum of the age-specific person selection probabilities (S) was larger than 2, then the selection probability was ratio adjusted by a multiplicative adjustment factor so that all probabilities were scaled down to sum to exactly 2. Now, Brewer's method sets the pairwise selection probabilities at

$$P_{h(ij)} = \left[\frac{P_{h(i)}P_{h(j)}}{K}\right]\left[\frac{1}{1-P_{h(i)}} + \frac{1}{1-P_{h(j)}}\right] \tag{2.1}$$

by setting K at

$$K = 2 + \sum \frac{P_{h(i)}}{1-P_{h(i)}}, \tag{2.2}$$

where $i = i^{th}$ person in household h (whose selection probability depends on his or her age category: 1, 2, 3, 4, or 5) and

$j = j^{th}$ person in household h (whose selection probability depends on his or her age category: 1, 2, 3, 4, or 5),

where age category 1 corresponds to persons aged 12 to 17, 2 to persons aged 18 to 25, 3 to persons aged 26 to 34, 4 to persons aged 35 to 49, and 5 to persons aged 50 or older.

The sum of the pairwise selection probabilities taken over all unique pairs will be guaranteed to be exactly 1.

$$\sum_{i}\sum_{j>i} P_{h(ij)} = 1 \tag{2.3}$$

It also guarantees that the sum of the pairwise selection probabilities for an individual is equal to the individual's selection probability

$$\sum_{j \neq i} P_{h(ij)} = P_{h(i)} \tag{2.4}$$

for all values of i.

Note the above scheme always selects a pair of two eligible persons.

2.1.2 Case II: DUs with $S < 2$

If the sum S of person-level selection probabilities was less than 2, the method used in survey years 1991 to 2001 consisted of dividing $2 - S$ equally among the three dummy persons

added to the household, and then used Brewer's method to select a pair, as in Case I. However, if the household had two or more persons, we preferred a pair of real persons to have a greater chance of being selected. To achieve this goal, the individual selection probabilities, $P_{h(i)}$, were scaled upward by the factor F_s such that their sum came close to but did not exceed 2 and such that each person selection probability did not exceed the maximum allowed probability of 0.99. Thus, denoting the revised person selection probabilities by $P'_{h(i)}$, the factor F_s is given by

$$F_s = Min\left\{\frac{T(\lambda)}{S}, \frac{0.99}{\max\left\{P_{h(i)}\right\}}\right\}, \tag{2.5}$$

where $T(\lambda) = S + \lambda(2 - S)$ and λ is set to 0.5. Note that if λ is chosen as 0, then $F_s = 1$ and the selection scheme would follow that of Case I. The individual person probabilities are scaled upward by the factor F_s so they either sum to 2 or sum as close to 2 as possible. Denote S' as the sum of the selection probability after scale adjustment by F_s. If S' is exactly 2, then dummy persons are not needed. If S' is less than 2, then three dummy persons are added to the DU.

Now, for Brewer's method, we set the pairwise selection probabilities similar to (2.1), as

$$P'_{h(ij)} = \left[\frac{P'_{h(i)}P'_{h(j)}}{K'}\right]\left[\frac{1}{1-P'_{h(i)}} + \frac{1}{1-P'_{h(j)}}\right] \tag{2.6}$$

by setting K' at

$$K' = 2 + \sum_i \frac{P'_{h(i)}}{1-P'_{h(i)}}, \tag{2.7}$$

where $P'_{h(i)}$ and $P'_{h(j)}$ are the selection probabilities adjusted by the scaling factor F_s,

where $i = i^{th}$ person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5),

$j = j^{th}$ person in the household (whose selection probability depends on his or her age category: 0, 1, 2, 3, 4, or 5), and

where age category 0 corresponds to dummy persons, and categories 1 to 5 are defined as in Case I.

Note that we now have $\sum_{j \neq i} P'_{h(ij)} = P'_{h(i)}$. To maintain the original person selection probabilities despite the scale adjustment by F_s, we modified Brewer's method as follows. First, draw a random number, R, from a uniform (0,1) distribution. If $R \leq 1/F_s$, then select a pair using Brewer's method based on formula (2.6). However, if $R > 1/F_s$, then no persons are selected from

the household. In this way, the probability for selecting a pair (i,j) in household h becomes $P^*_{h(ij)} = P'_{h(ij)}/F_s$, which, in turn, gives the original person selection probabilities, $P_{h(i)}$. Unlike Case I, where a pair of eligible persons was always selected, this adjusted selection scheme allows for zero, one, or two persons to be selected from a DU.

2.2 Questionnaire Dwelling Unit Selection Probability

A dwelling unit was considered a selected QDU if it had completed the screening interview and had at least one person selected for the questionnaire interview. QDUs with at least one respondent were considered respondent QDUs.

The QDU selection probability was defined as

$$P_h^* = \left(1 - P_{h(00)}^*\right), \tag{2.8}$$

where $P^*_{h(00)}$ is the probability of not selecting any person. For the DUs with an unadjusted sum of age-specific selection probabilities larger than or equal to 2 (Case I), $P^*_{h(00)}$ is 0. It follows from Section 2.1, under Case II, $P^*_{h(00)}$ can be calculated as

$$P_{h(00)}^* = \left(1 - \frac{1}{F_s}\right) + \frac{3}{F_s}\left[\frac{P'_{h(0)}P'_{h(0)}}{K'}\right]\left[\frac{1}{1-P'_{h(0)}} + \frac{1}{1-P'_{h(0)}}\right], \tag{2.9}$$

where $P'_{h(0)}$ is the selection probability of a dummy person when person selection probabilities are adjusted by F_s.

Table 2.1 Building Blocks of the QDU and Person Pair Samples: Dwelling Units and Persons in the 2008–2012 NSDUHs

	2008			2009			2010			2011			2012		
Domain	**Sel.[1]**	**Resp.[2]**	**% Rate[3]**	**Sel.[1]**	**Resp.[2]**	**% Rate[3]**	**Sel.[1]**	**Resp.[2]**	**% Rate[3]**	**Sel.[1]**	**Resp.[2]**	**% Rate[3]**	**Sel.[1]**	**Resp.[2]**	**% Rate[3]**
DUs															
Total DUs Screened	160,114	142,159	88.79	161,377	142,933	88.57	166,532	147,010	88.28	179,293	156,048	87.04	178,586	153,873	86.16
QDUs															
Total QDUs	58,942	48,180	81.74	58,288	48,088	82.50	58,702	48,113	81.96	61,441	50,133	81.60	60,621	48,850	80.58
Persons															
Total Persons	85,711	67,928	79.25	84,785	68,007	80.21	84,997	67,804	79.77	88,536	70,109	79.19	87,656	68,309	77.93
12-17	26,228	22,250	84.83	26,157	22,398	85.63	25,908	21,960	84.76	27,911	23,510	84.23	27,147	22,473	82.78
18-25	28,793	22,875	79.45	28,158	22,686	80.57	28,164	22,793	80.93	28,589	22,876	80.02	28,639	22,529	78.67
26-34	8,337	6,560	78.69	8,242	6,591	79.97	8,545	6,780	79.34	8,323	6,543	78.61	8,304	6,484	78.08
35-49	12,995	9,674	74.44	12,855	9,616	74.80	12,979	9,668	74.49	12,220	9,149	74.87	12,364	9,076	73.41
50+	9,358	6,569	70.20	9,373	6,716	71.65	9,401	6,603	70.24	11,493	8,031	69.88	11,202	7,747	69.16
Pairs															
Total Pairs[4]	26,769	19,748	73.77	26,497	19,919	75.17	26,295	19,691	74.88	27,095	19,976	73.73	27,035	19,459	71.98
0,0[5]	83,217	N/A	N/A	84,645	N/A	N/A	88,308	N/A	N/A	94,607	N/A	N/A	93,252	N/A	N/A
0, 12-17	8,634	8,094	93.75	8,432	7,936	94.12	8,595	7,906	91.98	9,402	8,651	92.01	9,017	8,277	91.79
0, 18-25	9,932	9,213	92.76	9,870	9,081	92.01	10,093	9,270	91.85	10,306	9,497	92.15	10,325	9,461	91.63
0, 26-34	3,771	3,303	87.59	3,798	3,318	87.36	3,914	3,418	87.33	3,930	3,386	86.16	3,856	3,327	86.28
0, 35-49	4,601	3,835	83.35	4,565	3,810	83.46	4,659	3,843	82.49	4,431	3,704	83.59	4,368	3,645	83.45
0, 50+	5,235	4,013	76.66	5,126	4,042	78.85	5,146	3,998	77.69	6,277	4,919	78.37	6,020	4,681	77.76
12-17, 12-17	4,319	3,654	84.60	4,288	3,648	85.07	4,160	3,525	84.74	4,649	3,885	83.57	4,507	3,668	81.38
12-17, 18-25	3,449	2,668	77.36	3,595	2,852	79.33	3,444	2,718	78.92	3,756	2,921	77.77	3,627	2,759	76.07
12-17, 26-34	897	690	76.92	872	724	83.03	922	752	81.56	834	685	82.13	825	658	79.76
12-17, 35-49	3,944	3,012	76.37	3,979	3,061	76.93	3,948	3,044	77.10	3,855	2,918	75.69	3,813	2,812	73.75
12-17, 50+	666	482	72.37	703	532	75.68	679	493	72.61	766	565	73.76	851	631	74.15
18-25, 18-25	5,872	4,261	72.56	5,588	4,168	74.59	5,502	4,165	75.70	5,476	4,015	73.32	5,476	3,901	71.24
18-25, 26-34	1,103	821	74.43	1,102	820	74.41	1,140	851	74.65	1,049	768	73.21	1,079	794	73.59
18-25, 35-49	1,706	1,112	65.18	1,576	1,059	67.20	1,639	1,098	66.99	1,469	994	67.67	1,582	1,053	66.56
18-25, 50+	859	546	63.56	839	547	65.20	844	537	63.63	1,057	666	63.01	1,074	660	61.45
26-34, 26-34	890	614	68.99	886	635	71.67	903	621	68.77	858	599	69.81	880	604	68.64
26-34, 35-49	545	380	69.72	447	312	69.80	512	354	69.14	492	314	63.82	469	320	68.23
26-34, 50+	241	145	60.17	251	149	59.36	251	162	64.54	302	192	63.58	315	177	56.19
35-49, 35-49	873	535	61.28	917	569	62.05	886	542	61.17	748	474	63.37	833	487	58.46
35-49, 50+	453	269	59.38	454	239	52.64	449	245	54.57	477	271	56.81	466	272	58.37
50+, 50+	952	559	58.72	1,000	604	60.40	1,016	584	57.48	1,307	709	54.25	1,238	663	53.55

DU = dwelling unit; N/A = not applicable; QDU = questionnaire dwelling unit.

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

[1] Selected pairs are based on the screener age.
[2] Respondent pairs are based on the questionnaire age and are comprised only of respondent persons.
[3] These rates are unweighted and based only on the total selected and total responding counts of pairs.
[4] Total pairs excludes dummy person pairs.
[5] Among the completed screening dwelling units, no person was selected in this dwelling unit.

3. Brief Description of the Generalized Exponential Model

In survey practice, design-based weights are typically adjusted in three steps: (1) for extreme values (ev) via winsorization, (2) for nonresponse (nr) via weighting classes, and (3) for poststratification (ps) via raking ratio adjustments. If weights are not treated for extreme values, the resulting estimates, although unbiased, will tend to have low precision. The bias introduced by winsorization is alleviated to some extent through ps. The nr adjustment is a correction for bias introduced in estimates based only on responding units, and ps is an adjustment for coverage (typically undercoverage) bias and variance reduction due to correlation between the study and control (usually demographic) variables.

There are limitations in the existing methods of weight adjustment for ev, nr, and ps. It would be desirable to adjust for bias introduced in the ev step (when extreme weights are treated via winsorization) in that the sample distribution for various demographic characteristics is preserved. For the nr step, there are general raking type methods, such as the scaled constrained exponential model developed by Folsom and Witt (1994), where the lower and upper bounds can be suitably chosen by use of a separate scaling factor. The factor is set as the inverse of the overall response propensity. It would be desirable to have a model for the nr adjustment factor so that the desired lower and upper bounds on the factor are part of the model. Note that the lower bound on the nr adjustment factor should be one, as it is interpreted as the inverse of the probability of response for a particular unit. For the ps step, on the other hand, the general calibration methods of Deville and Särndal (1992), such as the logit method, allow for built-in lower (L) and upper (U) bounds (for ps, typically L < 1 < U). However, it would be desirable to have nonuniform bounds (L_k, U_k) depending on the unit *k* such that the final adjusted weight, w_k, could be controlled within certain limits. An important application of this feature would be weight adjustments in the presence of ev to allow some control on the final adjustment of the initially identified extreme values.

A modification of the earlier method of the scaled constrained exponential model of Folsom and Witt (1994), termed as the method of the generalized exponential model (GEM) and proposed by Folsom and Singh (2000), provides a unified approach to the three weight adjustments for ev, nr, and ps, and it has the desired features mentioned above. The functional form of the GEM adjustment factor is provided in Appendix A. It generalizes the logit model of Deville and Särndal (1992), typically used for ps, such that the bounds (L, U) may depend on *k*. Thus, it provides a built-in control on ev during both ps and nr adjustments. In addition, the bounds are internal to the model and can be set to chosen values (e.g., $L_k = 1$ in the nr step). If there is a low frequency of ev in the final ps, then a separate ev step may not be necessary.

In fitting GEM to a particular problem, the choice of a large number of predictor variables along with tight bounds will have an impact on the resulting unequal weighting effect (UWE) and the proportion of extreme values. In practice, this leads to somewhat subjective considerations of trade-off between the target set of bounds for a given set of factor effects and the target UWE and the target proportion of extreme values. It also may be beneficial to look at

the proportion of "outwinsors" (a term coined to signify the extent of residual weights after winsorization), which is probably more realistic in determining the robustness of estimates in the presence of extreme values.

A large increase in the number of predictor variables in GEM typically would result in a higher UWE, thus indicating a possible loss in precision. This was checked by comparing SUDAAN-based standard errors of a key set of estimates computed from two sets of calibration models, one baseline using only the main effects and the other using the final model. The results are presented in Chapter 7.

To implement GEM, several steps need to be followed: (1) define and create all the covariates; (2) define the extreme weights; (3) fit the GEM model. The details of practical aspects of GEM implementation can be found in Chapters 4 and 5 of this report and Chapter 4 of Chen et al. (2014).

4. Predictor Variables for the Questionnaire Dwelling Unit and Pair Weight Calibration via the Generalized Exponential Model

We note that unlike the person-level weight calibration, the control totals for the questionnaire dwelling unit (QDU)-level and person pair-level poststratification are not available from the U.S. Census Bureau. A way around this problem is to take advantage of the two-phase nature of the design, in which the screener data provides a large sample containing demographic information that can be used to derive control totals for the QDU-level and person pair-level sampling weight calibrations, as well as for the selected person poststratification adjustment. The stability of control totals from the screener dwelling unit (SDU)-level data can be improved by poststratification of the SDU sample using person-level counts from the census. This was indeed done and is documented in the person-level weight calibration report (Chen et al., 2014).

4.1 Questionnaire Dwelling Unit Weight Calibration

After the nonresponse and poststratification adjustments at the SDU level, which are common to the person-level weight calibration, the QDU sample weights were adjusted in three steps: poststratification of selected QDUs, nonresponse adjustment of respondent QDUs, and poststratification of respondent QDUs. The set of initially proposed predictor variables for these adjustments using generalized exponential model (GEM) were set to be common and to correspond to those used for the SDU nonresponse and poststratification adjustments. The variables are of two types: Those used for SDU nonresponse adjustment are 0/1 indicators, while those used for SDU poststratification adjustment are counting variables. The variables of the first type (0/1 indicators) are population density, group quarters, race/ethnicity of householder, percentage of persons in segment who are black or African American, percentage of persons in segment who are Hispanic or Latino, percentage of owner-occupied dwelling units (DUs) in segment, segment-combined median rent and housing value, and household type. Variables of the second type (counting variables) represent the number of eligible persons within each DU who fall into the various demographic categories of race, age group, Hispanicity, and gender. Note that the State and quarter variables are represented as both binary and counting variables. Thus, not only are DU counts within a specific State or quarter in the QDU sample controlled to the corresponding totals obtained from the SDU sample, but also counts of persons living in the DUs in the QDU sample are controlled to totals from the SDU sample. These person-level totals match the census estimates because of the SDU-level poststratification to census counts. It may be noted that in the poststratification of selected QDUs and the nonresponse adjustment of the respondent QDUs steps, demographic information from screener data was used in defining covariates, whereas in the poststratification of the selected QDUs step, questionnaire demographic information was used.

Exhibit 4.1 lists all predictor variables proposed for QDU-level calibration and identifies them as counting, binary, or both. Various main effects and higher level factor effects based on

the predictor variables were included in the GEM modeling. As stated previously, all adjustment steps at the QDU level used a common set of proposed predictor variables.

4.2 Pair Weight Calibration

Like QDU, the initial set of weight components in pair weight calibration are the same as the set obtained from the SDU-level weight calibration. The SDU-calibrated weight is multiplied by the pair-level design weight, which in turn was adjusted in four steps: poststratification of selected pairs, nonresponse adjustment of respondent pairs, poststratification of respondent pairs, and the extreme weight adjustment of respondent pairs. All the adjustment steps for pair weights utilized the same set of initially proposed predictor variables, which included a subset of those used for the person-level nonresponse adjustment. This included segment characteristic variables, such as population density, percentage of persons in segment who are black or African American, percentage of persons in segment who are Hispanic or Latino, percentage of owner-occupied DUs in segment, and segment-combined median rent and housing value. Also included were pair-specific covariates, such as the demographic characteristics of pair age, pair race/ethnicity, and pair gender, as well as dwelling unit characteristics, such as race/ethnicity of householder, household type, household size, and group quarters indicators. State and quarter indicators were included as well. However, for two-factor effects, instead of individual State, State/region was used due to insufficient sample size. This resulted in a 12-level variable where the eight large sample States were kept separate, and the remainder of States were grouped according to the four census regions. All variables were defined as 0/1 indicators. These proposed predictor variables and their levels are shown in Exhibit 4.2.

In the poststratification of selected pairs and the nonresponse adjustment of respondent pairs, screener data were used in the definition of the pair-specific variables such as pair age, pair race/ethnicity, and pair gender, whereas in the poststratification and extreme weight adjustment of respondent pairs, these variables were obtained from the questionnaire. For the latter case, in addition to the variables described above, indicator covariates corresponding to selected pair domains were included to perform Hajek-type ratio adjustments via weight calibration, as mentioned in Chapter 1. The selected pair domains were limited to 10 of the 14 pair domains listed in Chapter 1. (Parent-child pairs where the child was in the 15- to 17-year-old age range and sibling-sibling-younger sibling focus pairs were not included in the poststratification.) The inclusion of these pair domain covariates led to the use of two sets of control totals in the modeling. Details of the construction of these control totals can be found in Appendix B.

Exhibit 4.1 Definitions of Levels for QDU-Level Calibration Modeling Variables

Age[b]
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1]

Gender[b]
1: Male, 2: Female[1]

Group Quarter Indicator[a]
1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Hispanicity[b]
1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]

Household Size[b]
Continuous Variable Count of Individuals Rostered with DU

Household Type (Ages of Persons Rostered within DU)[a]
1: 12-17, 18-25, 26+; 2: 12-17, 18-25; 3: 12-17, 26+; 4: 18-25, 26+; 5: 12-17, 6: 18-25; 7: 26+[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)[a]
1: 50-100%,[1] 2: 10-<50%, 3: 0-<10%

Percentage of Segments That Are Black or African American[a]
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino[a]
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Population Density[a]
1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter[a,b]
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race (3 Levels)[b]
1: White,[1] 2: Black or African American, 3: Other

Race (5 Levels)[b]
1: White,[1] 2: Black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: Two or More Races

Race/Ethnicity of Householder[a]
1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

Relation to Householder[a]
1: Householder or Spouse, 2: Child, 3: Other Relative, 4: Nonrelative[1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[a,2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

States[a,b,3]
Model Group 1: 1: Connecticut, 2: Maine, 3: Massachusetts,[1] 4: New Hampshire, 5: New Jersey, 6: New York, 7: Pennsylvania, 8: Rhode Island, 9: Vermont
Model Group 2: 1: Illinois, 2: Indiana, 3: Iowa, 4: Kansas, 5: Michigan, 6: Minnesota, 7: Missouri, 8: Nebraska, 9: North Dakota, 10: Ohio, 11: South Dakota, 12: Wisconsin[1]
Model Group 3: 1: Alabama, 2: Arkansas, 3: Delaware, 4: District of Columbia, 5: Florida, 6: Georgia, 7: Kentucky, 8: Louisiana, 9: Maryland, 10: Mississippi, 11: North Carolina,[1] 12: Oklahoma, 13: South Carolina, 14: Tennessee, 15: Texas, 16: Virginia, 17: West Virginia
Model Group 4: 1: Alaska, 2: Arizona,[1] 3: California, 4: Colorado, 5: Idaho, 6: Hawaii, 7: Montana, 8: Nevada, 9: New Mexico, 10: Oregon, 11: Utah, 12: Washington, 13: Wyoming

State/Region[a,3]
Model Group 1: 1: New York, 2: Pennsylvania, 3: Other[1]
Model Group 2: 1: Illinois, 2: Michigan, 3: Ohio, 4: Other[1]
Model Group 3: 1: Florida, 2: Texas, 3: Other[1]
Model Group 4: 1: California, 2: Other[1]

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.
[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.
[2] Segment-Combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percentage owner-occupied.
[3] The States or district assigned to a particular model is based on census regions.
[a] Binary variable.
[b] Counting variable.

Exhibit 4.2 Definitions of Levels for Pair-Level Calibration Modeling Variables

Group Quarter Indicator
1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Household Size
1: DU with 2 Persons,[1] 2: DU with 3 Persons, 3: DU with ≥ 4 Persons

Pair Age (15 Levels)
1: 12-17 and 12-17,[1] 2: 12-17 and 18-25, 3: 12-17 and 26-34, 4: 12-17 and 35-49, 5: 12-17 and 50+, 6: 18-25 and 18-25, 7: 18-25 and 26-34, 8: 18-25 and 35-49, 9: 18-25 and 50+, 10: 26-34 and 26-34, 11: 26-34 and 35-49, 12: 26-34 and 50+, 13: 35-49 and 35-49, 14: 35-49 and 50+, 15: 50+ and 50+

Pair Age (6 Levels)
1: 12-17 and 12-17,[1] 2: 12-17 and 18-25, 3: 12-17 and 26+, 4: 18-25 and 18-25, 5: 18-25 and 26+, 6: 26+ and 26+

Pair Age (3 Levels)
1: 12-17 and 12-17,[1] 2: 12-17 and 18+, 3: 18+ and 18+

Pair Gender
1: Male and Female,[1] 2: Female and Female, 3: Male and Male

Pair Race/Ethnicity (10 Levels)
1: White and White,[1] 2: White and Black or African American, 3: White and Hispanic or Latino, 4: White and Other, 5: Black or African American and Black or African American, 6: Black or African American and Hispanic or Latino, 7: Black or African American and Other, 8: Hispanic or Latino and Hispanic or Latino, 9: Hispanic or Latino and Other, 10: Other and Other

Pair Race/Ethnicity (5 Levels)
1: Two or More Races Pair, 2: Hispanic or Latino Pair, 3: Black or African-American Pair, 4: White Pair,[1] 5: Other Pair

Pair Race/Ethnicity (4 Levels)
1: Two or More Races Pair or Other and Other, 2: Hispanic or Latino Pair, 3: Black or African-American Pair, 4: White Pair[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)
1: 50-100%,[1] 2: 10-<50%, 3: 0-<10%

Percentage of Segments That Are Black or African American
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

Population Density
1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race/Ethnicity of Householder
1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

Exhibit 4.2 Definitions of Levels for Pair-Level Calibration Modeling Variables (continued)

State/Region

Model Group 1:1: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, Rhode Island, Vermont; 2: Alabama, Arkansas, Delaware, District of Columbia, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Virginia, West Virginia;[1] 3: New York; 4: Pennsylvania; 5: Florida; 6: Texas

Model Group 2: 1: Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota, Wisconsin;[1] 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming; 3: Michigan; 4: Illinois; 5: Ohio; 6: California

States[3]

Model Group 1: 1: Alabama, 2: Arkansas, 3: Connecticut, 4: Delaware, 5: District of Columbia, 6: Florida, 7: Georgia, 8: Kentucky, 9: Louisiana, 10: Maine, 11: Maryland,[1] 12: Massachusetts, 13: Mississippi, 14: New Hampshire, 15: New Jersey, 16: New York, 17: North Carolina, 18: Oklahoma, 19: Pennsylvania, 20: Rhode Island, 21: South Carolina, 22: Tennessee, 23: Texas, 24: Vermont, 25: Virginia, 26: West Virginia

Model Group 2: 1: Alaska, 2: Arizona,[1] 3: California, 4: Colorado, 5: Idaho, 6: Illinois, 7: Indiana, 8: Iowa, 9: Hawaii, 10: Kansas, 11: Michigan, 12: Minnesota, 13: Missouri, 14: Montana, 15: Nebraska, 16: Nevada, 17: New Mexico, 18: North Dakota, 19: Ohio, 20: Oregon, 21: South Dakota, 22: Utah, 23: Washington, 24: Wisconsin, 25: Wyoming

Pair Relationship Associated with Multiplicity

1: Parent-Child (12-14)*
2: Parent-Child (12-17)*
3: Parent-Child (12-10)*
4: Parent*-Child (12-14)
5: Parent*-Child (12-17)
6: Parent*-Child (12-20)
7: Sibling (12-14)-Sibling (15-17)
8: Sibling (12-17)-Sibling (18-25)
9: Spouse-Spouse/Partner-Partner
10: Spouse-Spouse/Partner-Partner with Children (younger than 18)

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-Combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percentage owner-occupied.

[3] The States or district assigned to a particular model is based on combined census regions.

* The pair member focused on.

5. Definition of Extreme Weights

An important feature of the generalized exponential model (GEM) is the built-in provision of extreme value (ev) treatment. Sampling weights are often classified as extreme (high or low) if they fall outside the interval, median ± 3 × interquartile range (IQR). The interval is set for prespecified domains defined usually by design variables corresponding to deep stratification.[7] Similar to previous NSDUHs, for the GEM modeling used in the 2012 NSDUH, a more conservative (narrower) interval was defined, median ± 2.5 × IQR. The narrower interval better prevents the adjusted weights from crossing the standard interval boundaries by treating weights near but not outside the commonly used boundaries (i.e., those that have the most potential to become extreme) as extreme as well.

Denote the interval boundaries (or critical values) for low and high extreme values by $b_{k(l)}$ and $b_{k(u)}$, respectively. For implementing extreme value control via GEM, the variable m_k was defined as the minimum of $b_{k(u)} / w_k$ and one for high extreme weights, and the maximum of $b_{k(l)} / w_k$ and one for low extreme weights, where w_k represents the sampling weight before adjustment, and $b_{k(u)}$ and $b_{k(l)}$ denote the critical values for the extreme weights. Note that under this definition, for high extreme weights, the more extreme the weight is, the smaller m_k will be, and, conversely, for low extreme weights, the more extreme the weight is, the bigger m_k will be. Nonextreme weights had a value of one for m_k. The upper and lower bounds for the adjustment factors were defined, respectively, as the product of m_k and the upper and lower boundary parameters of GEM. GEM allows inputs of up to three different upper and lower boundary parameters (L_1 and U_1, L_2 and U_2, L_3 and U_3) for high, non-, and low extreme weights. By applying a small upper boundary parameter for high extreme weights and a large lower boundary parameter for low extreme weights, the extreme weights can be controlled in the modeling process.

5.1 Questionnaire Dwelling Unit Extreme Weight Definition

For the questionnaire dwelling unit (QDU)-level weight adjustment, extreme weights were defined using a nested hierarchy of six domains:

1. State;
2. State sampling region;
3. State by household type;

Levels of household type indicate whether the household has members who are youths, young adults, or adults, where youth signifies 12- to 17-year-olds, young adult 18- to 25-year-olds, and adult 26 years or older.

[7] Deep stratification refers to the stratification that was used in the sample design. In the case of the 2012 survey, deep stratification refers to the cross-classification of State sampling region by age group.

a. Youth, Young Adult, Adult;
b. Youth, Young Adult;
c. Youth, Adult;
d. Young Adult, Adult;
e. Youth Only;
f. Young Adult Only; and
g. Adult Only.

4. State sampling region by household type;

5. State by household type by household size (1, 2, 3, 4+); and

6. State sampling region by household type by household size.

The hierarchy is used to satisfy the minimum of 30 observations for defining the boundaries for extreme values. If this sample size requirement is not met at the lower level, then the next level up in the hierarchy is used.

5.2 Person Pair Extreme Weight Definition

The pair selection probability is a function of the selection probability of each person in the pair given by formula (2.1) or (2.6), depending on the sum of the person selection probabilities within the household as discussed in Section 2.1. This probability can be very small if the selection probabilities of individual members are small. For example, consider a particular selected dwelling unit (DU) from the 2012 survey. This DU gave rise to a selected pair of respondents, both aged 50 or older. The selection probability in this DU was 0.10207 for a respondent aged 50 or older. Using the formula (2.6) in Chapter 2, the pair selection probability was computed to be 0.000227082. Therefore, the inverse of the selection probability, the pair-level design weight, was 4403.70. Thus, a small pair selection probability can create a high initial weight, which is the product of the screener dwelling unit (SDU) weight and the person pair design-based weight.

As mentioned in the introduction, it turns out to be difficult to select suitable domains for defining extreme weights for pair-level data. However, as was done for the 1999–2012 surveys, the extreme weight definition was based on the following hierarchy of domains:

1. Pair age group[8] (with three age categories, 12 to 25, 26 to 49, and 50+) by number (0, 1, 2+) of persons aged 12 to 25 in the household;

2. State cluster (with five levels [explained below]) by pair age group by number (0, 1, 2+) of persons aged 12 to 25 in the household;

3. State cluster (with three levels [explained below]) by pair age group by number (0, 1, 2+) of persons aged 12 to 25 in the household; and

4. State by pair age group by number of persons aged 12 to 25 (0, 1, 2+) in the household.

[8] Pair age in this case should not be confused with the modeling term, which has a finer level breakdown.

The hierarchy was used to satisfy the minimum of 30 observations for defining the boundaries for extreme values. If this sample size requirement was not met at the lower level, then the next level up in the hierarchy was used.

We now briefly introduce the considerations behind the above definition for extreme weight domains. The sample design prespecified the person-level selection probability within State by five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, 50+). Age groups 12 to 17 and 18 to 25 have a relatively similar selection probability, and the same is true for age groups 26 to 34 and 35 to 49. The 50+ group, however, has a quite different selection probability from the other groups. Furthermore, since the 12 to 17 and 18 to 25 age groups have large selection probabilities, they have a very high chance of being selected if the household has persons in these age groups. Therefore, the number of persons aged 12 to 25 in the household has a significant impact on the type of pair selected and the pair selection probability. Taking into consideration these design-related features, a suitable domain to define the pair-level extreme weight seems to be given by State by pair age group by number of persons aged 12 to 25 in the household.

The hierarchy of domains mentioned above was used to satisfy the minimum of 30 observations. However, it was found that for many ev domains the minimum sample size requirement was not met. To alleviate this problem, States were grouped into a small number of clusters, such as three or five. The assignment of States to clusters was determined by the clustering algorithm in PROC CLUSTER in SAS, where the clustering variable was defined as the average person-level weight (ANALWT) for each of the five age groups within each State. The choice of the average person-level weight for each group for each State was motivated from the objective of finding a single variable that would reflect the design-based difference in pair selection probabilities across States. Even with clustering of States, the ev domain sample size was insufficient in some cases, so the most general level of the hierarchy, the national level, was required. Furthermore, at the national level, we had to collapse some pair age categories in forming domains of reasonable sample size to define extreme weights. More specifically, for the national level, we collapsed all levels of number of persons aged 12 to 25 for the pair age groups of 50+, 50+ and 26 to 49, 50+. In addition, levels 1 and 2+ of number of persons aged 12 to 25 were combined for the pair age group of 26 to 49, 26 to 49.

6. Weight Calibration at Questionnaire Dwelling Unit and Pair Levels

The 2012 National Survey on Drug Use and Health (NSDUH) was based on probability sampling so that valid inferences can be made from survey findings about the target population. Probability sampling refers to sampling in which every unit on the frame is given a known, nonzero probability for inclusion in the survey. This is required for unbiased estimation of the population total. The assumption of nonzero inclusion probability for every pair of units in the frame also is required for unbiased variance estimation. The basic sampling plan involved four stages of selection across two phases of design: within Phase I, (1) the selection of census tracts within each State sampling (SS) region, (2) the selection of subareas or segments (comprised of U.S. Census Bureau blocks) within SS regions; (3) the selection of dwelling units (DUs) within these subareas; and, finally, within Phase II, (4) the selection of eligible individuals within DUs. Specific details of the sample design and selection procedures for the sample can be found in the 2012 NSDUH sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

As part of the postsurvey data-processing activities, analysis weights that reflected the selection probabilities from various stages of the sample design were calculated for respondents. These sample weights were adjusted at the DU (screening sample), questionnaire dwelling unit (QDU), person, and paired respondent levels (the latter three all based around the drug questionnaire sample) to account for bias due to extreme values (ev), nonresponse (nr), and coverage.

The final sample weights for Phase I screener dwelling units (SDU) and Phase II QDU, person, and pair levels for the 2012 samples consisted of products of several factors, each representing either a probability of selection at some particular stage or some form of ev, nr, or ps calibration adjustment. In the following sections, we describe the QDU and pair weight components in greater detail. In summary, the first 10 factors were defined for all SDUs and reflected the fully adjusted SDU sample weight. The remaining components branched to reflect QDU and pair selection probabilities, as well as additional adjustments for ev, nr, and ps. Note that the final QDU and pair weights for the 2012 survey sample are the product of all weight components for each type of sample, illustrated in Exhibits 6.1 and 6.2.

For QDU data, generalized exponential modeling (GEM) calibration modeling was applied by partitioning the data into four groups of States: Northeast, South, Midwest, and West, based on census regions in the interest of computational feasibility. Previous experience showed that with current computing power, the large number of variables and records prevented any further reduction of modeling groups.

For pair data, GEM modeling was initially applied by partitioning the pair data into four groups based on census regions. However, there were not enough observations in each group to fit a comprehensive model to reduce bias. Alternatively, a single model was attempted for the whole pair data, but it was rejected as not practical due to computational limitations. A compromise approach was adopted by combining census regions into two groups: Northeast with South and Midwest with West. This grouping proved both manageable and desirable as it

assisted in bias reduction, ease of modeling, and workload reduction. Exhibit 6.3 provides more details of the data partition for GEM modeling. The resulting sample sizes of selected and respondent units for the pair and QDU data partitions are shown for the 2008–2012 surveys in Table 6.1.

It may be noted that for the pair data in the 1999, 2000, and 2001 surveys, the built-in ev control feature of GEM was not used until the final respondent pair ev adjustment step. The reason for this is that the definition for ev domain was not finalized before the pair data calibration process was begun. However, for the 2002–2012 survey pair data, the built-in ev control feature was used for each adjustment step.

Exhibit 6.1 Summary of 2012 NSDUH QDU Sample Weight Components

Phase I Screener Dwelling Unit Level

Design Weight Components	
#1	Inverse Probability of Selecting Census Tract
#2	Inverse Probability of Selecting Segment
#3	Quarter Segment Weight Adjustment
#4	Subsegmentation Inflation Adjustment
#5	Inverse Probability of Selecting SDU
#6	Subsampling of Added SDU Adjustment
#7	SDU Release Adjustment
Weight Adjustment*	
#8	SDU Nonresponse Adjustment ***(res.sdu.nr)***
#9	SDU Poststratification Adjustment ***(res.sdu.ps)***
#10	SDU Extreme Value Adjustment ***(res.sdu.ev)***

Phase II Questionnaire Dwelling Unit Level

Design Weight Component	
#11	Inverse of Selection Probability of at Least One Person in the Dwelling Unit
Weight Adjustment*	
#12	Selected QDU Poststratification to SDU-Based Control Totals ***(sel.qdu.ps)***
#13	Respondent QDU Nonresponse Adjustment ***(res.qdu.nr)***
#14	Respondent QDU Poststratification to SDU-Based Control Totals ***(res.qdu.ps)***
#15	Respondent QDU Extreme Value Adjustment ***(res.qdu.ev)***

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

* These adjustments use the generalized exponential model (GEM), which also involves pre- and post-processing in addition to running the GEM macro. See Exhibit 4.1 (Chen et al., 2014). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 6.3.

Exhibit 6.2 Summary of 2012 NSDUH Person Pair Sample Weight Components

Phase I Screener Dwelling Unit Level

Design Weight Components	
#1	Inverse Probability of Selecting Census Tract
#2	Inverse Probability of Selecting Segment
#3	Quarter Segment Weight Adjustment
#4	Subsegmentation Inflation Adjustment
#5	Inverse Probability of Selecting SDU
#6	Subsampling of Added SDU Adjustment
#7	SDU Release Adjustment
Weight Adjustment*	
#8	SDU Nonresponse Adjustment ***(res.sdu.nr)***
#9	SDU Poststratification Adjustment ***(res.sdu.ps)***
#10	SDU Extreme Value Adjustment ***(res.sdu.ev)***

Phase II Person Pair Level

Design Weight Component	
#11	Inverse of Selection Probability of a Person Pair in SDU
Weight Adjustment*	
#12	Selected Pair Poststratification to SDU-Based Control Totals ***(sel.pr.ps)***
#13	Respondent Pair Nonresponse Adjustment ***(res.pr.nr)***
#14	Respondent Pair Poststratification Adjustment to SDU-Based Control Totals ***(res.per.ps)***
#15	Respondent Pair Extreme Value Adjustment ***(res.per.ev)***

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

* These adjustments use the generalized exponential model (GEM), which also involves pre- and post-processing in addition to running the GEM macro. See Exhibit 4.1 (Chen et al., 2014). For computational feasibility, all weight adjustments were done using the four model groups based on census regions defined in Exhibit 6.3.

Exhibit 6.3 U.S. Census Bureau Regions/Model Groups

Model Group	Census Region
QDU	
1	**Northeast (9 States)** Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont
2	**Midwest (12 States)** Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
3	**South (16 States and the District of Columbia)** Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia
4	**West (13 States)** Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming
Pair	
1	**Northeast + South (25 States and the District of Columbia)** Alabama, Arkansas, Connecticut, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maine, Massachusetts, Maryland, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, West Virginia
2	**Midwest + West (25 States)** Alaska, Arizona, California, Colorado, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oregon, South Dakota, Utah, Washington, Wisconsin, Wyoming

Table 6.1 Sample Size, by Model Group at QDU and Pair Levels

Model Group	2008		2009		2010		2011		2012	
	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs	Selected QDUs	Completed QDUs
QDU										
Northeast	11,490	9,216	11,605	9,340	11,627	9,339	11,997	9,456	12,616	9,917
South	17,703	14,747	17,756	14,909	17,880	14,857	19,690	16,487	18,345	15,019
Midwest	16,763	13,650	16,382	13,491	16,670	13,686	17,045	13,752	16,984	13,687
West	12,986	10,567	12,545	10,348	12,525	10,231	12,709	10,438	12,676	10,227
Total	58,942	48,180	58,288	48,088	58,702	48,113	61,441	50,133	60,621	48,850
Model Group	**2008**		**2009**		**2010**		**2011**		**2012**	
	Selected Pairs	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**	**Selected Pairs**	**Completed Pairs**
Pair										
Northeast + South	13,060	9,700	13,058	9,806	12,872	9,590	13,686	10,127	13,619	9,723
Midwest + West	13,709	10,048	13,439	10,113	13,423	10,101	13,409	9,849	13,416	9,736
Total	26,769	19,748	26,497	19,919	26,295	19,691	27,095	19,976	27,035	19,459

QDU = questionnaire dwelling unit.

Note: The 2008–2010 sample sizes reflect the removal of falsified cases found in Pennsylvania and Maryland. The 2011 and 2012 samples were not affected. For additional information, see Section B.3.5 in Appendix B of the 2011 and 2012 national findings reports (Center for Behavioral Health Statistics and Quality, 2012; 2013).

6.1 Phase I SDU-Level Weight Components

A total of 10 weight components for the SDU level correspond to selection probabilities and nonresponse, poststratification, and extreme value adjustment factors. Note that this differs from previous NHSDAs and NSDUHs in that a new design-based component was incorporated at the beginning of the process so that corresponding weight component numbers are incremented by one when compared to previous survey years with an otherwise similar weighting scheme. The first seven components in the Phase I sample weights reflect the probability of selecting the DUs. These components were derived from (1) the probability of selecting the census tract within each State sampling (SS) region, (2) the probability of selecting the geographic segment within each SS region, (3) a quarter segment weight adjustment, (4) a subsegmentation inflation factor, (5) the probability of selecting a DU from within each counted and listed sampled segment, (6) the probability of inclusion of added DUs, and (7) DU percent release adjustment. The three remaining weight components, #8 through #10, are GEM calibration adjustments accounting for (8) DU nonresponse at the screening level, (9) DU poststratification to census controls, and (10) DU-level extreme value adjustment, although in 2009 extreme value adjustment at this stage was deemed unnecessary, and thus Weight Component #10 was set to one for all respondent DUs. The person-level, QDU-level, and person pair-level weights use the product of the above 10 weight components as the common initial weight before further adjustments. For more detailed information on Weight Components #1, #2, and #4 through #7, refer to the 2012 NSDUH sample design report (Morton et al., 2013), and for more detail on Weight Components #3 and #8 through #10, see the 2012 person-level sampling weight calibration report (Chen et al., 2014).

Note that from 2008 to 2012, there was an occasional second subsegmentation step when the initial partitioning of segments was insufficient due to out-of-date census counts or the segment was still too large to list after the original subsegmentation. This second partitioning was not accounted for in the weighting over these survey years. A comparison was done to evaluate the effect of this omission, and it was determined that the missing second subsegmenting factor in the analysis weight had minimal impact on estimates. Therefore, weights for these years were not reproduced. Additional detail can be found in the 2012 NSDUH sample design report (Morton et al., 2013).

6.2 QDU Weight Components

6.2.1 QDU Weight Component #11: Inverse of Selection Probability of at Least One Person in the Dwelling Unit

The selection of a questionnaire dwelling unit from all completed SDUs is based on the outcome of a variant of Brewer's method, which may select zero, one, or two persons. Any pair of survey eligible residents within the dwelling unit had some known, nonzero chance of being selected for the survey. The value for Weight Component #11 is equal to the inverse of the probability that at least one person in the dwelling unit is selected (see Section 2.2 for details).

6.2.2 QDU Weight Component #12: Selected QDU Poststratification to SDU-Based Control Totals

This poststratification factor adjusts the weights for selected QDUs to the SDU-based control totals. The SDU-based control totals are obtained by using the calibrated SDU weights. This adjustment step provides more stable controls for the subsequent nonresponse adjustment (Weight Component #13). Exhibit 4.1 lists the initially proposed variables for GEM modeling. The predictor variables are either 0/1 indicators or counting variables representing the number of persons who fall into a given demographic domain. The counting variables are derived from the screener demographic information. It may be noted that during screening, the only required demographic information was the age of each person rostered. Thus, other demographic information necessary for weight calibration, such as race/ethnicity and gender may be missing for certain rostered eligible persons, and so imputation was done to replace this missing data. For more details on the imputation of screener demographic information, see Chen et al. (2014).

The details on the predictor variables retained in the model and model summary statistics can be found in Appendix C.

6.2.3 QDU Weight Component #13: Respondent QDU Nonresponse Adjustment

This nonresponse adjustment step accounts for the failure to obtain respondent person(s) from each and every selected QDU. The same set of initially proposed predictor variables were used as for the previous adjustment (#12).

See Appendix C for more details on the predictor variables retained in the model and model summary statistics.

6.2.4 QDU Weight Component #14: Respondent QDU Poststratification to SDU-Based Control Totals

This final poststratification for all respondent QDUs utilized the same set of initially proposed predictor variables as previous adjustments. The corresponding control totals were obtained from the SDU-level sample, as was done for Weight Component #12.

See Appendix C for more details on the predictor variables retained in the model and model summary statistics.

6.2.5 QDU Weight Component #15: Respondent QDU Extreme Value Adjustment

The extreme weight proportions for the final poststratified weights were acceptably low, eliminating the need for the extreme value adjustment. Weight Component #15 was set to one for each responding QDU.

6.3 Pair-Level Weight Components

Exhibit 4.2 lists the initially proposed predictor variables for the following adjustment steps via GEM.

6.3.1 Pair Weight Component #11: Inverse of Selection Probability of a Person Pair in the Dwelling Unit

Selection of pairs of individuals from all eligible persons residing within the dwelling unit is based on the outcome of a variant of Brewer's method, which may select zero, one, or two persons. Any pair of survey eligible residents within the dwelling unit has some known, nonzero chance of being selected for the survey. When two persons are selected, a pair is formed. The pair selection probability is determined by either formula (2.1) or formula (2.6) in Chapter 2. This weight component is the inverse of the selection probability discussed above.

6.3.2 Pair Weight Component #12: Selected Pair Poststratification to SDU-Based Control Totals

Similar to QDU Weight Component #12, this step was motivated by the consideration that the larger sample of all possible pairs provides more stable control totals for the respondent pair nonresponse adjustment. The weights of selected pairs were poststratified to the control totals that derived from calibrated SDU weights of all possible pairs. The pair-level demographic variables for all selected pairs, such as pair age group, pair race/ethnicity, etc., were derived from screener demographic information.

The details on the predictor variables retained in the model and model summary statistics can be found in Appendix H.

6.3.3 Pair Weight Component #13: Respondent Pair Nonresponse Adjustment

If both persons in the selected pair completed interviews successfully, the pair then was considered a respondent pair. This adjustment step accounts for failure to obtain respondent pairs

from all selected pairs. In this step, respondent pair weights were adjusted to the control totals based on the full sample of selected pairs. Due to the low response rate of person pairs, this step had a relatively large adjustment on the weights. The same set of proposed predictor variables was used as for Weight Component #12. Similar to Weight Component #12, the pair-level demographic variables for all selected pairs, such as pair age group, pair race/ethnicity, etc., were derived from screener demographic information.

See Appendix H for more details on the predictor variables retained in the model and model summary statistics.

6.3.4 Pair Weight Component #14: Respondent Pair Poststratification to SDU-Based Control Totals

This final poststratification utilized the same set of initially proposed predictor variables as previous adjustment steps. In addition, 10 pair relationship domain-level indicator variables were added to the set of covariates. The control totals for GEM calibration were derived from the SDU sample of all possible pairs of eligible persons, as was done for Weight Component #12. The calibration control totals for these 10 domains used household-level person counts and the final QDU weights. As mentioned in the introduction, use of these household-level count totals for pair relationship domains in GEM calibration provided Hajek-type weight adjustment in the interest of obtaining more stable estimates. In setting up calibration covariates, multiplicity factors were needed. These factors, as discussed in the introduction, are used in constructing estimates for person-level parameters based on pair-related drug behavior. The factors depend on the pair domains of interest. For a selected set of pair domains, multiplicity factors are provided along with the pair-level analysis weights. See Chapter 11 in the editing and imputation report (Carpenter et al., 2014) for more detail on the creation of and imputation of missing values in the pair relationship, multiplicity, and household-level person counts. See Chapter 4 for more detail on the use of multiplicities and household-level person counts in poststratification.

Unlike Weight Components #12 and #13, demographic covariates were based on data from the questionnaire instead of information pulled from the dwelling unit screener.

For more details on the predictor variables retained in the GEM model and model summary statistics, see Appendix H.

6.3.5 Pair Weight Component #15: Respondent Pair Extreme Weight Adjustment

We checked the extreme weight proportions for the weights up to Weight Component #14, using the extreme weight domains (see Section 5.2). Even though the previous adjustment steps utilized the built-in extreme weight control feature of GEM, the extreme weight proportions were still high enough to cause concern that they might produce unreliable estimates. Therefore, the extreme weight adjustment via GEM was implemented, using the same final set of predictor variables kept in the model for Weight Component #15. This step was successful in reducing the extreme weight proportion in all model groups. For details, see Appendix J.

7. Evaluation of Calibration Weights

During the weight calibration process, several criteria for quality control were implemented to assess model adequacy. In this chapter, we describe the individual procedures and a summary of their results. All tables referred to in this chapter can be found in Appendices D through G and I through L.

7.1 Response Rates

Table D.1 in Appendix D displays the final selected and responding questionnaire dwelling unit (QDU) sample sizes from the 2012 National Survey on Drug Use and Health (NSDUH) for various national domains. This table also shows the weighted response rates. Most domains reflect the overall 75.87 percent response rate, with most rates relatively close to 80 percent, although the highest response rate is 93.20 percent, from the Group level of the Group Quarters category. The lowest response rate came from Race/Ethnicity of Householder Non-Hispanic or Latino Other, with 72.15 percent.

Table I.1 in Appendix I displays the final selected and responding pair-level sample sizes from the 2012 survey for various national domains. Due to the nature of the pair data, the response rates were lower in all domains examined than at the QDU level, with an overall response rate of 63.76 percent. The response rates range from a low of 37.40 percent in the pair race/ethnicity Black or African American and Other category to a high of 81.51 percent from the Pair Age Group 12-17,12-17. This extreme range of response rates is probably due to a combination of small sample sizes and response burden as a result of selection of pairs within households among various domains. Like at the QDU level, the top response rates are among the younger respondents (as measured by household type for the QDU data and pair age for the pair data). This pattern may be related to the relatively high response rates in the group level of the variable group quarters since it includes college dormitories.

7.2 Proportions of Extreme Values and Outwinsors

During the stages of modeling adjustments (i.e., nonresponse [nr] and poststratification [ps]), one major issue of concern when deciding the adequacy of a particular model was the extent of the resulting proportions of extreme value (ev) and outwinsor weights (see Sections 5.1 and 5.2 for these definitions). For each weight adjustment step, these proportions are computed before and after the step for various domains. Prior to adjustment, the product of all weight components is used to compute proportions of evs and outwinsors, while after the adjustment the product includes the new adjustment factor. If the proportion of evs and outwinsors are deemed high, a separate ev treatment step after ps could be performed. This was done for the pair-level weights. Details of this step are explained in Section 6.3.5. A separate ev treatment step was deemed unnecessary for the QDU-level weights.

Tables E.1 and E.2 and Tables J.1 through J.3 present percentages of evs at the QDU level and the pair level, respectively, for various domains. Unweighted percentages are the percentage of actual counts of units defined as evs relative to the total sample size. Weighted percentages reflect the percentage of total ev weights relative to the total sample weight, while

outwinsor percentages represent the total amount of residual weight when the weights are trimmed to the critical values (used for ev definition) relative to the total sample weight. For evaluation purposes, the outwinsor percentage is considered the most important of the three percentages, as this gave a measure of the impact of winsorization (or trimming) of ev weights (if we performed this treatment). See Sections 5.1 and 5.2 for the domains that were used to define extreme values.

7.3 Slippage Rates

The slippage rate for a given domain is defined as the relative percentage difference between the sampling weights and the external control totals, both before and after ps. The control totals for QDU and person pair ps are derived from the screener dwelling unit (SDU) weights, which were poststratified to U.S. Census Bureau population estimates (Chen et al., 2014). Table F.1 displays QDU national domain-specific weight sums for both before and after ps, as well as the desired totals to be met through ps. Table K.1 shows the same for the pair sample. These tables also show the relative percentage difference, or the amount of adjustment necessary (positive or negative) to meet the desired totals. The first relative difference is used explicitly during the ps modeling procedure to identify potential problems for convergence. Large differences in domains with relatively small sample sizes are indicative of potential large adjustment factors, which may cause problems in convergence while satisfying bound constraints. The reason is that adjustments required for one domain may have an adverse effect on another domain when a unit belongs to both.

As an example, consider that Table F.1, for the 2012 QDU domain household size of two, indicates a sample size of 16,974 with a total design-based weight of 54,585,381 and a census total of 54,556,667 with an initial slippage rate of -0.07 percent, which would imply a common weight adjustment approximately equal to 1.000681, if this were the only calibration control. Similarly, looking at pair data in Table K.1, the pair domain category of pair age 18-25, 18-25 has a sample size of 3,901, a design-based weight of 12,270,700, and a census total of 12,545,793, showing an initial slippage of -2.69 percent. The resultant required adjustment would be approximately equal to 1.027695, if this were the only control. However, in the generalized exponential model (GEM), all controls are simultaneously satisfied under a complex algorithm that allows for different adjustment factors for different units.

7.4 Weight Adjustment Summary Statistics

Tables G.1, G.2, and L.1 through L.3 display summary statistics on the product of weight components before and after all stages of adjustment for the QDU and person pair, respectively. The summary statistics include sample size (*n*), minimum (min), maximum (max), median (med), 25th percentile (Q1), 75th percentile (Q3), and the unequal weighting effect (UWE). Note that in Tables L.2 and L.3 the sample size for pair age group, pair race/ethnicity, and pair gender are slightly different. This is because those variables were defined using screening demographic information in the nonresponse adjustment of respondent pairs, while in the poststratification of respondent pairs, they were defined from questionnaire demographic information. Because UWE is directly affected by weight adjustment factors and extreme weights, these values—along with the percentage of extreme weights as noted in Section 7.2—were used as guidelines for determining model adequacy.

7.5 Sensitivity Analysis of Drug Use Estimates

It is known that, in general, there is a trade-off between bias reduction and variance reduction. For instance, with GEM (for nr or ps), enlarging a simple model (such as the one with only main effects) has the potential of further reducing the bias. At the same time, this enlargement also may be associated with a corresponding increase in the variance of the estimate due to additional variability caused by estimating the model parameters. To check for possible overfitting of the GEM model, we conducted a sensitivity analysis for respondent QDU poststratification for the QDU weights, respondent pair poststratification, and extreme weight adjustment for the person pair weights. A simple baseline model was fitted with the same bounds and maximum number of iterations as was used for the chosen (more complex) final model. We then looked for substantial changes in point estimates and standard errors (SEs). For the QDU weights, some household-level characteristics were selected such as family income, number of youths in the household, whether the household had health insurance coverage, and number of elders living in the household. The estimates and SEs are displayed in Table 7.1. For the person pair weights, selected licit and illicit drug use prevalence rates of 12- to 17-year-olds were calculated from parent-child pairs, and estimates and SEs of the estimates based on pair weights are shown in Tables 7.2a to 7.7b.

As seen in Table 7.1, the estimates and their SEs for the two models (baseline and the final) are generally similar to each other for the QDU weights. However, among the person pair estimates and SEs, there are some differences, but they do not seem significant in general.

Since the sensitivity analyses for both QDU- and pair-level calibrated weights seem to indicate that adding more covariates does not introduce an undesirable degree of instability in the estimates or their SEs, the final, more complex GEM models were deemed reasonable.

Table 7.1 Estimates of Totals and SEs for Domains of Interest Based on QDU Sample: 2012

Domain	2012				
	n	Baseline (B)[1]	Final (F)[2]	(B-F)/F% (Estimate)	(B-F)/F% (SE)
Households with Family Income					
$0 - < $10,000	5,078	10,460,715 (329,056)	10,444,004 (329,693)	0.16	-0.19
$10,000 - < $20,000	6,555	15,622,561 (398,493)	15,638,194 (400,681)	-0.10	-0.55
$20,000 - < $30,000	5,748	14,195,740 (359,770)	14,214,304 (360,747)	-0.13	-0.27
$30,000 - < $40,000	5,331	13,379,640 (327,509)	13,364,505 (327,477)	0.11	0.01
$40,000 - < $50,000	5,080	12,604,296 (332,618)	12,606,932 (331,778)	-0.02	0.25
$50,000 - < $75,000	7,703	19,169,185 (415,719)	19,159,643 (415,892)	0.05	-0.04
$75,000 - < $100,000	5,265	13,710,548 (383,622)	13,716,657 (385,188)	-0.04	-0.41
$100,000+	8,090	20,005,297 (507,343)	20,003,744 (507,248)	0.01	0.02
Households with Number of Youths (< 18)					
0	20,419	77,402,125 (1,069,348)	77,400,559 (1,071,586)	0.00	-0.21
1	11,659	17,372,185 (274,586)	17,379,411 (276,588)	-0.04	-0.72
2	9,719	15,069,648 (275,565)	15,063,588 (274,382)	0.04	0.43
3	4,603	6,422,304 (162,360)	6,404,101 (161,754)	0.28	0.37
4+	2,450	2,881,721 (96,317)	2,900,323 (99,755)	-0.64	-3.45
Households with Insurance Coverage					
Yes	39,384	99,535,440 (1,147,242)	99,551,648 (1,148,703)	-0.02	-0.13
No	9,466	19,612,543 (381,660)	19,596,335 (381,456)	0.08	0.05
Households with Number of Older Adults (65+)					
0	43,406	89,339,074 (983,199)	89,341,651 (986,485)	-0.00	-0.33
1	3,721	19,513,672 (525,590)	19,511,535 (525,200)	0.01	0.07
2	1,676	10,100,864 (375,736)	10,094,422 (376,016)	0.06	-0.07
3+	47	194,373 (42,011)	200,375 (44,298)	-3.00	-5.16

QDU = questionnaire dwelling unit; SE = standard error.
Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last step of calibration, res.qdu.ps, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.2a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, by Mother Use: 2012

		2012		
Drug	**Mother User**	***n***	**Baseline[1]**	**Final[2]**
Alcohol				
Lifetime	Yes	1,976	30.36 (1.74)	30.03 (1.77)
	No	288	22.64 (3.93)	22.65 (3.96)
	Overall	2,264	29.16 (1.61)	28.89 (1.63)
Past Year	Yes	1,636	26.27 (1.85)	26.01 (1.89)
	No	628	14.28 (2.06)	14.25 (2.07)
	Overall	2,264	22.73 (1.47)	22.52 (1.49)
Past Month	Yes	1,232	13.56 (1.59)	13.54 (1.68)
	No	1,032	9.24 (1.33)	9.25 (1.33)
	Overall	2,264	11.50 (1.06)	11.49 (1.09)
Cigarettes				
Lifetime	Yes	1,458	16.07 (1.51)	15.80 (1.51)
	No	806	10.55 (1.52)	10.57 (1.54)
	Overall	2,264	13.75 (1.14)	13.59 (1.14)
Past Year	Yes	597	15.79 (2.05)	15.34 (2.03)
	No	1,667	7.88 (1.05)	7.84 (1.05)
	Overall	2,264	9.63 (0.94)	9.49 (0.94)
Past Month	Yes	532	9.61 (1.79)	9.50 (1.79)
	No	1,732	4.43 (0.81)	4.29 (0.79)
	Overall	2,264	5.44 (0.75)	5.29 (0.73)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.2b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, by Father Use: 2012

		2012		
Drug	**Father User**	***n***	**Baseline[1]**	**Final[2]**
Alcohol				
Lifetime	Yes	1,385	31.37 (2.19)	31.58 (2.28)
	No	99	15.24 (4.48)	15.06 (4.48)
	Overall	1,484	30.19 (2.08)	30.30 (2.16)
Past Year	Yes	1,178	26.45 (2.22)	26.59 (2.33)
	No	306	11.82 (2.29)	11.44 (2.29)
	Overall	1,484	23.51 (1.85)	23.44 (1.92)
Past Month	Yes	971	15.55 (2.21)	15.62 (2.37)
	No	513	6.45 (1.28)	6.24 (1.25)
	Overall	1,484	12.38 (1.53)	12.32 (1.62)
Cigarettes				
Lifetime	Yes	1,059	20.53 (2.20)	20.29 (2.27)
	No	425	10.73 (2.76)	10.87 (3.02)
	Overall	1,484	17.11 (1.75)	16.96 (1.83)
Past Year	Yes	428	16.48 (3.02)	16.40 (3.14)
	No	1,056	10.97 (1.88)	10.95 (1.93)
	Overall	1,484	12.28 (1.60)	12.24 (1.64)
Past Month	Yes	377	13.01 (3.27)	13.55 (3.47)
	No	1,107	3.51 (0.89)	3.43 (0.89)
	Overall	1,484	5.41 (1.00)	5.43 (1.03)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.3a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Mother-Child (12 to 17) Pairs, by Mother Use: 2012

		2012		
Drug	**Mother User**	***n***	**Baseline[1]**	**Final[2]**
Any Illicit				
Lifetime	Yes	1,204	25.17 (2.08)	25.25 (2.09)
	No	1,060	15.88 (1.72)	16.01 (1.78)
	Overall	2,264	20.42 (1.36)	20.49 (1.38)
Past Year	Yes	229	21.82 (4.03)	21.65 (4.05)
	No	2,035	13.82 (1.19)	13.89 (1.21)
	Overall	2,264	14.48 (1.15)	14.51 (1.17)
Past Month	Yes	103	14.57 (4.77)	15.02 (5.03)
	No	2,161	6.98 (0.90)	7.01 (0.92)
	Overall	2,264	7.28 (0.89)	7.32 (0.91)
Marijuana				
Lifetime	Yes	1,080	17.40 (1.86)	17.45 (1.87)
	No	1,184	11.27 (1.40)	11.25 (1.41)
	Overall	2,264	13.92 (1.12)	13.89 (1.12)
Past Year	Yes	134	20.63 (5.11)	19.63 (4.91)
	No	2,130	10.60 (1.03)	10.60 (1.04)
	Overall	2,264	11.02 (1.01)	10.96 (1.01)
Past Month	Yes	68	12.38 (4.47)	12.18 (4.35)
	No	2,196	5.22 (0.75)	5.25 (0.76)
	Overall	2,264	5.38 (0.74)	5.41 (0.75)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.3b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug or Marijuana among Father-Child (12 to 17) Pairs, by Father Use: 2012

		2012		
Drug	**Father User**	***n***	**Baseline[1]**	**Final[2]**
Any Illicit				
Lifetime	Yes	898	25.51 (2.48)	26.17 (2.69)
	No	586	19.18 (2.94)	18.99 (2.95)
	Overall	1,484	22.98 (1.90)	23.28 (2.01)
Past Year	Yes	178	24.24 (6.13)	24.57 (6.55)
	No	1,306	15.95 (1.79)	16.02 (1.89)
	Overall	1,484	16.89 (1.74)	16.99 (1.84)
Past Month	Yes	116	23.46 (8.39)	23.69 (8.99)
	No	1,368	8.96 (1.35)	8.96 (1.43)
	Overall	1,484	10.07 (1.43)	10.09 (1.52)
Marijuana				
Lifetime	Yes	816	18.40 (2.32)	18.88 (2.55)
	No	668	10.93 (2.45)	10.93 (2.45)
	Overall	1,484	15.04 (1.68)	15.27 (1.79)
Past Year	Yes	135	25.14 (7.83)	25.84 (8.57)
	No	1,349	11.56 (1.63)	11.71 (1.73)
	Overall	1,484	12.59 (1.62)	12.79 (1.74)
Past Month	Yes	97	23.24 (10.27)	24.22 (11.35)
	No	1,387	6.98 (1.24)	7.05 (1.34)
	Overall	1,484	7.81 (1.32)	7.93 (1.44)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.4 Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Parent Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	Yes	2,227	30.76 (1.62)	30.52 (1.65)
	No	1,468	29.21 (2.26)	29.06 (2.29)
	Overall	3,695	30.12 (1.33)	29.92 (1.35)
Past Year	Yes	2,227	24.91 (1.47)	24.72 (1.51)
	No	1,468	21.28 (1.92)	20.94 (1.91)
	Overall	3,695	23.41 (1.18)	23.17 (1.20)
Past Month	Yes	2,227	12.92 (1.19)	12.88 (1.25)
	No	1,468	11.08 (1.52)	10.88 (1.48)
	Overall	3,695	12.16 (0.94)	12.06 (0.96)
Cigarettes				
Lifetime	Yes	2,227	14.91 (1.24)	14.75 (1.29)
	No	1,468	16.94 (1.86)	16.61 (1.83)
	Overall	3,695	15.75 (1.05)	15.52 (1.06)
Past Year	Yes	2,227	10.73 (1.09)	10.58 (1.11)
	No	1,468	11.97 (1.67)	11.76 (1.63)
	Overall	3,695	11.24 (0.94)	11.07 (0.94)
Past Month	Yes	2,227	5.56 (0.76)	5.48 (0.76)
	No	1,468	6.03 (1.08)	5.86 (1.05)
	Overall	3,695	5.75 (0.63)	5.64 (0.63)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.5 Percentages of Youths (12 to 17) Living with a Parent Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Parent-Child (12 to 17) Pairs, Asked Whether Their Parents Had Spoken to Them about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Parent Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	Yes	2,227	20.16 (1.35)	20.25 (1.41)
	No	1,468	24.77 (2.20)	24.87 (2.21)
	Overall	3,695	22.06 (1.23)	22.15 (1.25)
Past Year	Yes	2,227	15.29 (1.23)	15.28 (1.28)
	No	1,468	17.02 (1.93)	17.07 (1.92)
	Overall	3,695	16.00 (1.09)	16.01 (1.11)
Past Month	Yes	2,227	8.46 (1.00)	8.51 (1.06)
	No	1,468	8.64 (1.37)	8.65 (1.39)
	Overall	3,695	8.53 (0.83)	8.56 (0.87)
Marijuana				
Lifetime	Yes	2,227	14.16 (1.17)	14.23 (1.22)
	No	1,468	15.90 (1.81)	15.83 (1.79)
	Overall	3,695	14.88 (1.01)	14.89 (1.03)
Past Year	Yes	2,227	11.56 (1.11)	11.56 (1.17)
	No	1,468	12.65 (1.73)	12.60 (1.71)
	Overall	3,695	12.01 (0.96)	11.99 (0.98)
Past Month	Yes	2,227	6.42 (0.90)	6.49 (0.97)
	No	1,468	6.38 (1.09)	6.38 (1.10)
	Overall	3,695	6.40 (0.70)	6.45 (0.73)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.6a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Children about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Mother Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	0 times	171	23.88 (5.10)	24.15 (5.25)
	1-2 times	310	24.65 (3.85)	24.16 (3.91)
	A few times	546	25.42 (3.53)	24.67 (3.50)
	Many times	1,147	33.49 (2.25)	33.44 (2.30)
	Overall	2,174	29.36 (1.65)	29.14 (1.68)
Past Year	0 times	171	15.35 (3.74)	15.15 (3.78)
	1-2 times	310	18.83 (3.49)	18.70 (3.60)
	A few times	546	22.43 (3.49)	21.91 (3.45)
	Many times	1,147	26.16 (2.00)	26.07 (2.06)
	Overall	2,174	23.17 (1.53)	22.99 (1.56)
Past Month	0 times	171	7.53 (2.63)	7.77 (2.81)
	1-2 times	310	8.17 (2.15)	7.91 (2.07)
	A few times	546	10.60 (2.76)	10.15 (2.67)
	Many times	1,147	13.75 (1.54)	13.95 (1.65)
	Overall	2,174	11.58 (1.10)	11.58 (1.13)
Cigarettes				
Lifetime	0 times	171	11.53 (4.89)	10.97 (4.87)
	1-2 times	310	11.01 (2.48)	10.62 (2.38)
	A few times	546	11.05 (2.55)	11.14 (2.59)
	Many times	1,147	16.32 (1.51)	16.20 (1.52)
	Overall	2,174	13.83 (1.16)	13.70 (1.17)
Past Year	0 times	171	5.50 (2.14)	4.86 (2.00)
	1-2 times	310	7.70 (2.25)	7.27 (2.11)
	A few times	546	9.09 (2.50)	9.19 (2.53)
	Many times	1,147	11.52 (1.23)	11.44 (1.24)
	Overall	2,174	9.81 (0.97)	9.68 (0.97)
Past Month	0 times	171	4.12 (1.99)	3.62 (1.87)
	1-2 times	310	5.36 (1.96)	5.10 (1.85)
	A few times	546	5.14 (2.06)	5.00 (2.01)
	Many times	1,147	6.12 (0.94)	6.06 (0.94)
	Overall	2,174	5.59 (0.78)	5.45 (0.76)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.6b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Alcohol and Tobacco among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Father Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Alcohol				
Lifetime	0 times	174	20.01 (5.17)	21.59 (5.66)
	1-2 times	247	23.73 (4.69)	23.00 (4.75)
	A few times	459	34.16 (4.00)	34.70 (4.24)
	Many times	497	30.66 (3.24)	30.67 (3.30)
	Overall	1,377	29.29 (2.11)	29.53 (2.19)
Past Year	0 times	174	13.01 (4.46)	13.61 (4.84)
	1-2 times	247	21.30 (4.61)	20.68 (4.67)
	A few times	459	24.11 (3.49)	24.44 (3.79)
	Many times	497	24.52 (2.84)	24.34 (2.85)
	Overall	1,377	22.43 (1.83)	22.45 (1.91)
Past Month	0 times	174	3.04 (1.48)	2.91 (1.45)
	1-2 times	247	10.85 (2.56)	10.10 (2.45)
	A few times	459	12.73 (3.03)	13.32 (3.39)
	Many times	497	12.55 (2.37)	12.33 (2.43)
	Overall	1,377	11.18 (1.42)	11.16 (1.52)
Cigarettes				
Lifetime	0 times	174	13.46 (4.95)	14.89 (5.44)
	1-2 times	247	14.25 (4.31)	13.89 (4.37)
	A few times	459	16.12 (3.12)	16.25 (3.42)
	Many times	497	18.51 (2.61)	17.85 (2.61)
	Overall	1,377	16.31 (1.70)	16.23 (1.79)
Past Year	0 times	174	10.70 (4.82)	11.79 (5.33)
	1-2 times	247	10.94 (4.16)	11.04 (4.28)
	A few times	459	9.63 (2.41)	9.40 (2.35)
	Many times	497	14.42 (2.49)	13.92 (2.50)
	Overall	1,377	11.71 (1.55)	11.62 (1.57)
Past Month	0 times	174	6.61 (4.10)	7.29 (4.54)
	1-2 times	247	3.58 (1.15)	3.54 (1.16)
	A few times	459	4.13 (1.28)	4.13 (1.31)
	Many times	497	6.48 (1.89)	6.41 (1.89)
	Overall	1,377	5.16 (0.97)	5.21 (1.01)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.7a Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Mother-Child (12 to 17) Pairs, for Mother in the Pair, Asked Whether She Had Spoken to Her Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Mother Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	0 times	171	16.81 (5.15)	16.94 (5.22)
	1-2 times	310	19.66 (3.65)	19.03 (3.54)
	A few times	546	19.04 (2.99)	19.41 (3.12)
	Many times	1,147	22.10 (1.82)	22.29 (1.86)
	Overall	2,174	20.52 (1.39)	20.64 (1.42)
Past Year	0 times	171	5.67 (1.77)	5.74 (1.76)
	1-2 times	310	11.75 (2.64)	11.38 (2.57)
	A few times	546	13.90 (2.70)	13.97 (2.76)
	Many times	1,147	17.25 (1.71)	17.45 (1.75)
	Overall	2,174	14.55 (1.19)	14.65 (1.20)
Past Month	0 times	171	2.73 (1.28)	2.52 (1.13)
	1-2 times	310	6.27 (1.88)	6.20 (1.86)
	A few times	546	7.58 (2.39)	7.72 (2.44)
	Many times	1,147	8.34 (1.28)	8.44 (1.32)
	Overall	2,174	7.32 (0.93)	7.39 (0.95)
Marijuana				
Lifetime	0 times	171	11.81 (4.81)	11.89 (4.83)
	1-2 times	310	11.66 (2.72)	11.39 (2.66)
	A few times	546	13.64 (2.62)	13.78 (2.68)
	Many times	1,147	15.11 (1.46)	15.15 (1.49)
	Overall	2,174	13.93 (1.15)	13.96 (1.16)
Past Year	0 times	171	4.60 (1.62)	4.51 (1.53)
	1-2 times	310	9.29 (2.55)	9.03 (2.47)
	A few times	546	10.87 (2.49)	10.89 (2.52)
	Many times	1,147	12.66 (1.38)	12.69 (1.40)
	Overall	2,174	10.98 (1.03)	10.97 (1.03)
Past Month	0 times	171	1.42 (0.69)	1.41 (0.66)
	1-2 times	310	5.14 (1.82)	5.13 (1.80)
	A few times	546	5.13 (2.12)	5.25 (2.15)
	Many times	1,147	6.22 (0.94)	6.23 (0.97)
	Overall	2,174	5.36 (0.76)	5.40 (0.77)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

Table 7.7b Percentages of Youths (12 to 17) Reporting Lifetime, Past Year, and Past Month Use of Any Illicit Drug and Marijuana among Father-Child (12 to 17) Pairs, for Father in the Pair, Asked Whether He Had Spoken to His Child about the Dangers of Tobacco, Alcohol, or Drug Use within the Past 12 Months: 2012

Drug	Father Talked about Dangers with Child	2012		
		n	Baseline[1]	Final[2]
Any Illicit				
Lifetime	0 times	174	17.12 (4.72)	18.72 (5.16)
	1-2 times	247	21.40 (4.75)	21.62 (4.90)
	A few times	459	19.79 (3.31)	20.46 (3.62)
	Many times	497	27.18 (3.10)	26.96 (3.24)
	Overall	1,377	22.43 (1.91)	22.83 (2.03)
Past Year	0 times	174	13.47 (4.54)	15.02 (4.99)
	1-2 times	247	15.00 (4.36)	14.70 (4.45)
	A few times	459	15.20 (3.09)	15.84 (3.43)
	Many times	497	19.45 (2.67)	18.69 (2.65)
	Overall	1,377	16.48 (1.72)	16.56 (1.82)
Past Month	0 times	174	7.52 (3.71)	7.71 (3.71)
	1-2 times	247	5.29 (1.57)	4.99 (1.48)
	A few times	459	9.20 (2.61)	9.74 (3.02)
	Many times	497	12.46 (2.30)	11.90 (2.24)
	Overall	1,377	9.43 (1.33)	9.39 (1.42)
Marijuana				
Lifetime	0 times	174	6.97 (3.05)	8.19 (3.79)
	1-2 times	247	14.49 (4.33)	14.92 (4.50)
	A few times	459	14.38 (3.09)	14.96 (3.43)
	Many times	497	16.33 (2.54)	15.99 (2.58)
	Overall	1,377	14.24 (1.63)	14.54 (1.75)
Past Year	0 times	174	6.55 (3.04)	7.76 (3.78)
	1-2 times	247	12.46 (4.28)	12.28 (4.35)
	A few times	459	12.65 (3.05)	13.35 (3.40)
	Many times	497	13.15 (2.39)	12.65 (2.39)
	Overall	1,377	12.08 (1.58)	12.25 (1.69)
Past Month	0 times	174	2.57 (1.47)	2.62 (1.55)
	1-2 times	247	5.00 (1.54)	4.70 (1.46)
	A few times	459	8.32 (2.59)	8.90 (3.01)
	Many times	497	8.53 (2.03)	8.27 (2.01)
	Overall	1,377	7.10 (1.19)	7.16 (1.30)

Note: Standard errors of prevalence estimates are provided in parentheses.
[1] Baseline refers to the weight obtained from using a main effects only model for the last two steps of calibration, res.pr.ps and res.pr.ev, and a full model for preceding steps.
[2] Final refers to the weight obtained using a full model throughout all steps of calibration.

References

Carpenter, L., Edwards, S., Frechtel, P., Gulledge, K., Handley, W., King, S., Kroutil, L., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2014). Editing and imputation report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.107.006.007). Research Triangle Park, NC: RTI International.

Center for Behavioral Health Statistics and Quality. (2012). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 12-4713, NSDUH Series H-44). Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at http://www.samhsa.gov/data/NSDUH.aspx

Center for Behavioral Health Statistics and Quality. (2013). *Results from the 2012 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 13-4795, NSDUH Series H-46). Rockville, MD: Substance Abuse and Mental Health Services Administration. Available at http://www.samhsa.gov/data/NSDUH.aspx

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 41, RTI/0212800.001.107.004). Research Triangle Park, NC: RTI International.

Chen, P., Dai, L., Gordek, H., Grau, E., Shi, W., & Westlake, M. (2006). Person-level sampling weight calibration. In *2004 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-03-9028, Deliverable No. 26, RTI/0208726.174.003). Research Triangle Park, NC: RTI International.

Chen, P., Penne, M. A., & Singh, A. C. (2000). Experience with the generalized exponential model (GEM) for weight calibration for NHSDA. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 604-607). Alexandria, VA: American Statistical Association. Available at http://www.amstat.org/sections/srms/proceedings/

Chromy, J. R., & Singh, A. C. (2001). Estimation for person-pair drug-related characteristics in the presence of pair multiplicities and extreme sampling weights. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. Available at http://www.amstat.org/sections/srms/proceedings/

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376-382.

Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598-603). Alexandria, VA: American Statistical Association. Available at http://www.amstat.org/sections/srms/proceedings/

Folsom, R. E., & Witt, M. B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. In Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Toronto, Ontario, Canada (pp. 428-433). Alexandria, VA: American Statistical Association.

Frechtel, P., Warren, L. K., & Porter, J. (in press). *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI/0212800). Research Triangle Park, NC: RTI International.

Hajek, J. (1971). Comment on D. Basu's paper, "An essay on the logical foundations of survey sampling, part one" [pp. 203-234]. In V. P. Godambe, & D. A. Sprott (Eds.), *Foundations of statistical inference: A symposium* (p. 236). Toronto, Ontario, & Montreal, Quebec, Canada: Holt, Rinehart and Winston of Canada.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

Penne, M., Chen, P., & Singh, A.C. (2001) Person-pair sampling weight calibration using the generalized exponential model for the National Household Survey on Drug Abuse. In *Proceedings of the 2001 Joint Statistical Meetings*, *American Statistical Association, Survey Research Methods Section, Atlanta, GA* [CD-ROM]. Alexandria, VA: American Statistical Association. Available at http://www.amstat.org/sections/srms/proceedings/

Singh, A. C., & Mohl, C. A. (1996). Understanding calibration estimators in survey sampling. *Survey Methodology, 22*, 107-115.

Appendix A: Technical Details about the Generalized Exponential Model

Appendix A: Technical Details about the Generalized Exponential Model

A.1 Distance Function

Let $\Delta(w,d)$ denote the distance between the initial weights $d = \{d_k : k \in s\}$ and the adjusted weights *w*, with *k* being the kth unit in the sample and *s* being the sample selected. The distance function minimized under the generalized exponential model (GEM), subject to calibration constraints, is given by

$$\Delta(w,d) = \sum_{k\in s} \frac{d_k}{A_k} \left\{ (a_k - \ell_k) \log \frac{a_k - \ell_k}{c_k - \ell_k} + (u_k - a_k) \log \frac{u_k - a_k}{u_k - c_k} \right\}, \tag{A.1.1}$$

where $a_k = w_k / d_k, A_k = (u_k - \ell_k)/[(u_k - c_k)(c_k - \ell_k)]$ and ℓ_k, c_k, and u_k are prescribed real numbers. Let T_x denote the *p*-vector of control totals corresponding to predictor variables ($x_1, ..., x_p$). Then, the calibration constraints for the above minimization problem are

$$\sum_{k\in s} x_k d_k a_k = T_x. \tag{A.1.2}$$

The solution for the above minimization problem, if it exists, is given by a GEM with model parameters λ; that is,

$$a_k(\lambda) = \frac{\ell_k (u_k - c_k) + u_k (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}{(u_k - c_k) + (c_k - \ell_k) \exp\{A_k x_k' \lambda\}}. \tag{A.1.3}$$

Note that the number of parameters in the GEM should be $\leq n$, where *n* is the size of the sample *s*. This is also the dimension of vectors *d* and *w*. It follows from Equation A.1.3 that

$$\ell_k < a_k < u_k, k = 1, \ldots, n. \tag{A.1.4}$$

The usual raking ratio method (Singh & Mohl, 1996) of weight adjustment is a special case of the GEM, noting that for $\ell_k = 0$, $u_k = \infty$, $c_k = 1$, and $k = 1, \ldots, n$, we have

$$\Delta(w,d) = \sum_{k\in s} d_k a_k \log a_k - \sum_{k\in s} d_k (a_k - 1) \tag{A.1.5}$$

and $a_k(\lambda) = \exp(x_k' \lambda)$.

The logit method of Deville and Särndal (1992) is also a special case of the GEM, by setting $\ell_k = \ell$, $u_k = u$, and $c_k = 1$ for all *k*. The new method was introduced by Folsom and Singh (2000).

A.2 GEM Adjustments for Extreme Value Treatment, Nonresponse, and Poststratification

By choosing the user-specified parameters ℓ_k, c_k, and u_k appropriately, the unified GEM formula (A.1.3) can be justified for all three types of adjustment: extreme value treatment, nonresponse, and poststratification. For extreme value treatment via winsorization, denote the winsorized weights by $\{b_k\}$, where $b_k = d_k$ if d_k is not an extreme weight, and $b_k = \text{med}\{d_k\} \pm 3 * \text{IQR}$ if d_k is an extreme weight, where IQR denotes the interquartile range, and the median and quartiles for the weights are defined with respect to a suitable design-based stratum.

For the nonresponse adjustment, the sample is first divided into two parts: the nonextreme weight subsample and the extreme weight subsample. For nonextreme weights, the following are set: $\ell_2 = 1$, $c_2 = \rho^{-1}$, $u_2 = u > \rho^{-1}$, where ρ is the overall response propensity. For extreme weights with high weights, $\ell_k = \ell_1 m_k$, $c_k = \rho^{-1} m_k$, and $u_k = u_1 m_k$, where $m_k = b_k / d_k$ and $1 \le \ell_1 < \rho^{-1} = c_1 < u_1$ are prescribed numbers. Similarly, for extreme weights with low weights, $\ell_k = \ell_3 m_k$, $c_k = \rho^{-1} m_k$, $u_k = u_3 m_k$, and $1 \le \ell_3 < \rho^{-1} = c_3 < u_3$.

For the poststratification adjustment, the following weights are set: for nonextreme weights, $\ell_k = \ell_2$, $c_k = c_2 = 1$, and $u_k = u_2$; for high extreme weights, $\ell_k = \ell_1 m_k$, $c_k = m_k$, and $u_k = u_1 m_k$; and similarly, for low extreme weights, $\ell_k = \ell_3 m_k$, $c_k = m_k$, and $u_k = u_3 m_k$. The extreme value adjustment is identical to poststratifcation, except for tighter bounds on extreme weights resulting from the final poststratification.

Notice that the GEM allows the flexibility of specifying different bounds for different subsamples. In addition, the lower bound (in the case of nonresponse adjustments) can be made to equal one by choosing the center $c_k > 1$.

A.3 Newton-Raphson Steps

Let X denote the $n \times p$ matrix of predictor values, and for the vth iteration,

$$\Gamma_{\phi v} = \text{diag}\left(d_k \phi_k^{(v)}\right), \phi_k^{(o)} = 1,$$

where $\phi_k^{(v)} = \left[\left(u_k - a_k^{(v)}\right)\left(a_k^{(v)} - \ell_k\right)\right] \Big/ \left[\left(u_k - c_k\right)\left(c_k - \ell_k\right)\right]$.

Then, for the Newton-Raphson iteration v, the value of the p-vector λ is adjusted as

$$\lambda^{(v)} = \lambda^{(v-1)} + \left(X'\Gamma_{\phi,v-1}X\right)^{-1}\left(T_x - \hat{T}_x^{(v-1)}\right),$$

where $\lambda^{(0)} = 1$.

The convergence criterion is based on the Euclidean distance $\left\|T_x - \hat{T}_x^{(v)}\right\|$, which is defined as $\sqrt{\left(T_x - \hat{T}_x^{(v)}\right)'\left(T_x - \hat{T}_x^{(v)}\right)}$. At each iteration, it is checked to determine whether it is decreasing or not. If not, a half step[1] is used in the iteration increment.

A.4 Scaled Constrained Exponential Model

In National Household Surveys on Drug Abuse (NHSDAs)[2] prior to 1999, constrained exponential models (CEMs) were used for poststratification, and scaled CEMs were used for nonresponse adjustments. The CEM refers to the logit model of Deville and Särndal (1992), in which lower and upper bounds do not vary with k; that is, $\ell_k = \ell$, $u_k = u$, and $c_k = c = 1$, such that $\ell < 1 < u$. Thus, the CEM is a special case of the GEM. For the nonresponse adjustment, Folsom and Witt (1994) modified the CEM estimating equations by a scaling factor (ρ^{-1}, the inverse of the overall response propensity), such that $1 < \rho^{-1}a_k < \rho^{-1}u$. This implies that choosing ℓ in the CEM as ρ ensures that the scaled adjustment factor for nonresponse is at least one.

[1] A half step refers to halving the increment in the Newton-Raphson iterative process for fitting GEM.

[2] The National Household Survey on Drug Abuse (NHSDA) was renamed the National Survey on Drug Use and Health (NSDUH) in the 2002 survey year.

Appendix B: Derivation of Poststratification Control Totals

Appendix B: Derivation of Poststratification Control Totals

Unlike the person-level poststratification adjustment, the control totals for questionnaire dwelling unit (QDU)-level and person pair-level weight calibration could not be derived from the U.S. Bureau of the Census directly. Estimates of the number of households and person pairs were not available at the domains that we wanted to control, and person pair population estimates were not available even at a national level. However, by taking advantage of the two-phase design of the National Survey on Drug Use and Health (NSDUH), the screener dwelling unit (SDU) sample weights could be poststratified to census population estimates. The calibrated SDU weights then could be used as stable control totals for the QDU- and person pair-level sample weights. In addition to the SDU weights, the person pair-level weights were calibrated to a second set of controls derived from the questionnaire, called household-level person counts. These controls were applied to pairs that were members of the 10 selected pair domains given below.

1. Parent-child pairs, child aged 12 to 14, target population is parents whose children aged 12 to 14 live with them;
2. Parent-child pairs, child aged 12 to 14, target population is children aged 12 to 14 living with their parents;
3. Parent-child pairs, child aged 12 to 17, target population is parents whose children aged 12 to 17 live with them;
4. Parent-child pairs, child aged 12 to 17, target population is children aged 12 to 17 living with their parents;
5. Parent-child pairs, child aged 12 to 20, target population is parents whose children aged 12 to 20 live with them;
6. Parent-child pairs, child aged 12 to 20, target population is children aged 12 to 20 living with their parents;
7. Sibling-sibling pairs, older sibling aged 15 to 17, younger sibling aged 12 to 14, target population is siblings aged 15 to 17 whose siblings are aged 12 to 14;
8. Sibling-sibling pairs, older sibling aged 18 to 25, younger sibling aged 12 to 17, target population is siblings aged 18 to 25 whose siblings are aged 12 to 17;
9. Spouse-spouse and partner-partner pairs; and
10. Spouse-spouse and partner-partner pairs with children younger than the age of 18 living in the household.

B.1 Derivation of QDU-Level Poststratification Controls

The derivation of QDU-level poststratification controls was not directly possible. Instead, it had to be based on work done for the person-level calibration. At the person level, weights were calibrated to the control totals that we wished to reach. These weights then were altered in order to conform to use with QDU-level data.

B.1.1 Person Level

B.1.1.1 Receiving and Deriving Person-Level Poststratification Control Totals

Civilian, noninstitutionalized population estimates for ages 12 or older were provided by the Population Estimates Branch of the U.S. Bureau of the Census. We received two files, one at the national level and the other at the State level, each containing estimates of the population broken down by levels of month (12), Hispanicity (2), race (6), gender (2), and age (11).

The breakdown received from the census did not match the levels of the domains that we wanted to control. To account for this, we collapsed levels. From this altered data, we created datasets with model group-specific control totals. Observations in these datasets corresponded to a breakdown by quarter (4), Hispanicity (2), race (5), gender (2), age (11), and number of States[1] in the model group (number of States varied according to which census region was represented in the model group).

B.1.1.2 Adjusting SDU Data to the Control Totals

In the person-level weighting, the SDU weights were poststratified to meet control totals based on the population estimates received from the census. For NSDUH weighting, GEM was utilized to calibrate sample weights to multiple control totals. In doing so, each SDU received an adjustment factor, which, when multiplied by the initial weight, produced a final weight. The sum of all final weights corresponded to the civilian, noninstitutionalized population estimate for ages 12 or older, and the sum of all final weights in a domain corresponded to the control total for that domain. Note that there were a number of controls being calibrated to for each SDU, depending upon the domains to which the SDU belonged. The adjusted SDU weight reflected the civilian, noninstitutionalized population estimates for ages 12 or older and could be utilized as a basis for constructing controls at the QDU and person pair levels.

B.1.2 QDU Level

B.1.2.1 Deriving QDU-Level Poststratification Control Totals from Adjusted SDU Weights

Since there were no controls for QDU-level poststratification available directly, we used the adjusted SDU weights. For these weights to be applicable at the QDU level, the SDU-level data had to be restructured by sorting and summing over the domains to be used in the QDU-level calibration. This provided a dataset where the summed weight, which still added up to the proper population, was available for every domain to be utilized in the QDU calibration and thus could be used as a control total.

[1] The District of Columbia is included among States.

B.1.2.2 Adjusting QDU-Level Data to the Control Totals

As was done for the SDU data, the QDU-level data was adjusted via calibration in GEM of sample weights to multiple control totals. Each QDU received an adjustment factor, similar to that described for the SDU weight in B.1.1.2. The controls utilized in this calibration were based on the SDU weight as described in B.1.2.1 above. The adjusted weight was representative of the civilian, noninstitutionalized population estimates for ages 12 or older for all domains controlled within the modeling.

B.2 Derivation of Person Pair-Level Poststratification Controls

B.2.1 Deriving Person Pair-Level Poststratification Control Totals from Adjusted SDU Weights and Household-Level Person Counts

Analogous to the QDU weights, some of the person pair controls were based on the SDU weights. However, two sets of control totals were utilized in the modeling, with one set based on the SDU weights and the other set based on the questionnaire roster.

For most pair data domains—those other than the 10 pair domains based on relationship—the control totals for the poststratification adjustments were obtained from SDU data and were based on the number of possible pairs within SDUs. In order to obtain these pair counts belonging to various sociodemographic domains, the screener roster information was used to calculate all possible pairs within SDUs. For example, consider an SDU with two persons aged 12 to 17 and three persons aged 26 to 34. From this household composition, one can construct one pair of persons aged 12 to 17, three pairs of persons aged 26 to 34, and six pairs of persons aged 12 to 17 and 26 to 34. It follows that the total number of possible pairs in this SDU is 10, from which the number of pairs belonging to the domain of interest can be obtained.

On the other hand, for the 10 selected pair domains based on relationship, the control totals for the poststratification adjustments were obtained from the questionnaire roster. This involved calibrating the pair weights to the number of persons in households belonging to each domain of interest. These controls were obtained from the larger sample of singles and pairs (i.e., one or two persons selected from dwelling units) and were calculated at the QDU (household) level. The pair weights were adjusted by the appropriate multiplicity. See Chapter 11 in Carpenter et al. (2014) for details on the multiplicity counts and household-level control totals, which are referred to as household-level person counts.

B.2.2 Adjusting Person-Pair Level Data to the Control Totals

Like the SDU- and QDU-level data, the person pair-level data was adjusted via GEM. The use of two different types of controls required a minor modification to the GEM macro so that both sets of controls might be addressed simultaneously. Similar to the SDU- and QDU-level poststratification steps, each pair received an adjustment factor, which, when multiplied by the initial weight, produced a final weight. The sum of all final weights corresponded to the civilian, noninstitutionalized population estimate for ages 12 or older, and the sum of all final weights in a domain corresponded to the control total for that domain.

Appendix C: GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

Appendix C: GEM Modeling Summary for the Questionnaire Dwelling Unit Weights

This appendix summarizes each questionnaire dwelling unit (QDU) model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this appendix provides a model-specific overview of weight calibration, as opposed to a State- or domain-specific one.

For 2012, modeling involved taking four model groups through three adjustment steps: (1) selected dwelling unit poststratification, (2) respondent dwelling unit nonresponse adjustment, and (3) respondent dwelling unit poststratification. After the final poststratification, the adjusted sampling weights were reasonably distributed and did not require the additional treatment of the extreme value step.

Model-specific summary statistics are shown in Tables C.1a through C.4b. Included in these tables, for each stage of modeling, are the number of factor effects included; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; weighted, unweighted, and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. For more details on bounds, see Section 4.1. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the current weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data led to situations whereby inclusion of all originally proposed levels of covariates in the model was not possible. The text and exhibits in Sections C.1 through C.4 summarize the decisions made with regard to final covariates included in each model. For a list of the proposed initial covariates considered at each stage of modeling, see Exhibit C.2, and for the list of realized final model covariates, see Exhibits C.1.1 through C.4.3. The following sections establish a series of guidelines to assist in their interpretation.

C.1 Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 4.1 (included here as Exhibit C.1 for easy reference) in Chapter 4. There, a complete list is provided of all variables and associated levels used at any stage of modeling. Note that not all factors or levels were present in all stages of modeling, and the initial set of variables was the same across model groups but may change over stages of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment. Exhibits C.1.1 through C.4.3 provide lists of the proposed and realized covariates.

To help understand what effects were controlled for at each stage of the modeling, it was useful to create cross-classification tables as shown in Section C.3. Sections C.2 and C.3 explain how to use various exhibits for selected model variables to construct these tables.

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables

Age[b]
1: 12-17, 2: 18-25, 3: 26-34, 4: 35-49, 5: 50+[1]

Gender[b]
1: Male, 2: Female[1]

Group Quarter Indicator[a]
1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Hispanicity[b]
1: Hispanic or Latino, 2: Non-Hispanic or Latino[1]

Household Size[b]
Continuous Variable Count of Individuals Rostered with DU

Household Type (Ages of Persons Rostered within DU)[a]
1: 12-17, 18-25, 26+; 2: 12-17, 18-25; 3: 12-17, 26+; 4: 18-25, 26+; 5: 12-17; 6: 18-25; 7: 26+[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)[a]
1: 50-100%,[1] 2: 10-<50%, 3: 0-<10%

Percentage of Segments That Are Black or African American[a]
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino[a]
1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Population Density[a]
1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter[a,b]
1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race (3 Levels)[b]
1: White[1], 2: Black or African American, 3: Other

Race (5 Levels)[b]
1: White,[1] 2: Black or African American, 3: American Indian or Alaska Native, 4: Asian, 5: Two or More Races

Race/Ethnicity of Householder[a]
1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

Relation to Householder[a]
1: Householder or Spouse, 2: Child, 3: Other Relative, 4: Nonrelative[1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[a,2]
1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

Exhibit C.1 Definitions of Levels for QDU-Level Calibration Modeling Variables (continued)

State[a,b,3]
Model Group 1: 1: Connecticut, 2: Maine, 3: Massachusetts,[1] 4: New Hampshire, 5: New Jersey, 6: New York, 7: Pennsylvania, 8: Rhode Island, 9: Vermont
Model Group 2: 1: Illinois, 2: Indiana, 3: Iowa, 4: Kansas, 5: Michigan, 6: Minnesota, 7: Missouri, 8: Nebraska, 9: North Dakota, 10: Ohio, 11: South Dakota, 12: Wisconsin[1]
Model Group 3: 1: Alabama, 2: Arkansas, 3: Delaware, 4: District of Columbia, 5: Florida, 6: Georgia, 7: Kentucky, 8: Louisiana, 9: Maryland, 10: Mississippi, 11: North Carolina,[1] 12: Oklahoma, 13: South Carolina, 14: Tennessee, 15: Texas, 16: Virginia, 17: West Virginia
Model Group 4: 1: Alaska, 2: Arizona,[1] 3: California, 4: Colorado, 5: Idaho, 6: Hawaii, 7: Montana, 8: Nevada, 9: New Mexico, 10: Oregon, 11: Utah, 12: Washington, 13: Wyoming
State/Region[a,3]
Model Group 1: 1: New York, 2: Pennsylvania, 3: Other[1]
Model Group 2: 1: Illinois, 2: Michigan, 3: Ohio, 4: Other[1]
Model Group 3: 1: Florida, 2: Texas, 3: Other[1]
Model Group 4: 1: California, 2: Other[1]

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit.

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-Combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percentage owner-occupied.

[3] The States or district assigned to a particular model is based on census regions.

[a] Binary variable.

[b] Counting variable. A count of all persons in the household.

C.2 Glossary of Terms Used in the Description of the Variables in the Final Model

This glossary provides a list of general terms. Certain other terms are sometimes used within a particular section.

All levels present. All effects and all levels of the factor under consideration are in the model.

Coll. *(levels)*. Collapse these factor effects together. Factor effects that have been collapsed with others manifest themselves jointly in the model.

Conv. If model is not convergent, dropping or collapsing of variables is performed.

Do the same for *(effects)*. Repeat the previous step for all effect levels listed.

Drop all levels. All factor effects are completely removed from the model for all levels and any combinations involving this factor.

Drop *level(s)*. Collapse these factor effects into the reference set. The factor effects comprising the dropped levels are manifested jointly with either some or all of the factor effects in the reference set.

Drop *level(s)*; sing. During the modeling process the factor effects listed are removed from the model due to singularity.

Drop *level(s); zero cnts.* During the modeling process the factor effects listed are removed from the model due to zero sample.

Drop or Collapse using*. The asterisk is used as a wild card character to indicate all levels of the factor for that effect.

Factor effect. The factor effect represents the effects of levels considered for one factor, two factors, and higher order factors.

Hier. One or more of the factor effects in a higher order interaction is collapsed or dropped in an interaction at a lower order and the hierarchical effect carries up, either eliminating or combining factors of higher order interactions with that effect.

Reference/reference set. Factor effects composed of reference levels are not explicitly listed in the set of model variables. However, these effects manifest themselves either separately or in combination with other factors depending on the presence of other factors in the model.

C.3 How To Interpret Collapsing and Dropping of Factor Effects

To help visualize what effects are directly controlled for in our model, one can construct the table that reflects the collapsing scheme employed. The following is a complex example from the 2004 person-level modeling (Chen et al., 2006).

1. Locate the Factor Effect—Model 9 Person Nonresponse Adjustment.

Three-Factor Effects	Comments
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all levels of age in State (5).

2. Determine the initial range of possible levels for the variables by referring to the variable definitions. See Exhibits C.1 and H.1 for QDU- and pair-level variable definitions. In addition, the columns "Levels," "Proposed," and "Final" will provide counts of all factor effects, all explicitly proposed factors, and all explicitly controlled factors, but these are not necessary for construction of the cross-classification table. The following example is based upon person-level variables, but the process is the same.

State (for the model group in question, in this case, Model Group 9)
Model Group 9: 1: Alaska, 2: Hawaii, 3: Oregon, 4: Washington,[1] 5: California

Age
1: 12 to 17, 2: 18 to 25, 3: 26 to 34, 4: 35 to 49, 5: 50+[1]

Race (3 Levels)
1: White,[1] 2: Black or African American, 3: Other

3. Construct the cross-classification table.

For example, the initial proposed set of covariates in Race (4 Levels) is defined this way:

Race (4 Levels)	White	Black or African American	Asian	American Indian or Alaska Native

Shading indicates the reference-level set.

[1] This is the reference level for this variable. This is the level against which effects of other factor levels are measured.

This is the cross-classification table for the initial proposed set of covariates in State × Race (4 Levels):

State × Race (4 Levels)	White	Black or African American	Asian	American Indian or Alaska Native
AK				
HI				
OR				
WA				
CA				

Shading indicates the reference-level set.

The cross-classification table of interest for the initial proposed set of covariates in State × Age × Race (3 Levels) is as follows:

State × Age × Race (3 Levels)	White	Black or African American	Other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Shading indicates the reference-level set.

The number of respondents in the class State × Age × Race (3 Levels) at this stage of modeling would appear within each cell of the table. Construction of the other cross-

classification tables follows the same logic and is only necessary to the point of providing understanding of the final table.

4. Use the information under the "Comments" column definition to determine the combination of factors controlled.

One-Factor Effects	**Comments**
State	All levels present.
Race (4 Levels)	All levels present.
Age	All levels present.

Two-Factor Effects	**Comments**
State × Age	All levels present.
State × Race (4 Levels)	Coll. (1,3) & (1,4). Do the same for all other States except (2). Coll. (2,2), (2,3), & (2,4).
Age × Race (3 Levels)	All levels present.

The reason for the hier. instruction in the three-factor effect directions is the State × Race (4 Levels) interaction. It indicates a need to maintain the collapsing scheme when setting up any three-factor crosses involving State × Race. Following these directions, the resulting two-factor table we would then have to work with is as follows:

State × Race (4 Levels)	**White**	**Black or African American**	**Asian**	**American Indian or Alaska Native**
AK				
HI				
OR				
WA				
CA				

Shading indicates the reference-level set.

Returning to our instructions, we see that several other factor crosses have been affected by modeling:

Three-Factor Effects	**Comments**
State × Age × Race (3 Levels)	Coll. (2,1,2) & (2,1,3); hier. Repeat for all levels of age in State (2); hier. Coll. (1,4,2) & (1,4,3); conv. Drop (3,4,2); sing. Drop (3,*,*); conv. Coll. (5,1,2) & (5,1,3); conv. Repeat for all levels of age in State (5).

Construct the complete table, and then begin combining blocks as directed. The unshaded cells represent the factors directly controlled for by the model. The shaded cells represent the composite reference set, whose values may be obtained by utilizing the marginal sums, although when changes to the initially proposed set occur, it can make certain reference cell counts indistinguishable.

After following the directions, the resulting post-modeling cross-classification table should appear as follows:

State × Age × Race (3 Levels)	White	Black or African American	Other
AK × 12-17			
18-25			
26-34			
35-49			
50+			
HI × 12-17			
18-25			
26-34			
35-49			
50+			
OR × 12-17			
18-25			
26-34			
35-49			
50+			
WA × 12-17			
18-25			
26-34			
35-49			
50+			
CA × 12-17			
18-25			
26-34			
35-49			
50+			

Shading indicates the reference-level set.

Exhibit C.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights

Variables	Binary	Counting	Level	Proposed
One-Factor Effects		**76**	**76**	
Intercept	Yes		1	1
Population Density	Yes		4	3
Group Quarter	Yes		3	2
Race/Ethnicity of Householder	Yes		6	5
Rent/Housing	Yes		5	4
Segment % Black or African American	Yes		3	2
Segment % Hispanic or Latino	Yes		3	2
Segment % Owner-Occupied	Yes		3	2
Household Type	Yes		7	6
State	Yes	Yes	Model-specific	
Quarter	Yes	Yes	4	3
Age Group		Yes	5	4
Race		Yes	5	4
Hispanicity		Yes	2	1
Gender		Yes	2	1
Household Size		Yes	1	1
Two-Factor Effects				
Age × Race (3 Levels)		Yes	5 × 3	8
Age × Hispanicity		Yes	5 × 2	4
Age × Gender		Yes	5 × 2	4
Race (3 Levels) × Hispanicity		Yes	3 × 2	2
Race (3 Levels) × Gender		Yes	3 × 2	2
Hispanicity × Gender		Yes	2 × 2	1
State × Age		Yes	Model-specific	
State × Race (5 Levels)		Yes	Model-specific	
State × Gender		Yes	Model-specific	
State × Hispanicity		Yes	Model-specific	
% Black or African American × % Owner-Occupied	Yes		3 × 3	4
% Black or African American × Rent/Housing		Yes	3 × 5	8
% Hispanicity × % Owner-Occupied		Yes	3 × 3	4
% Hispanicity × Rent/Housing		Yes	3 × 5	8
% Owner × Rent/Housing	Yes		3 × 5	8
Three-Factor Effects				
Race (3 Levels) × Age × Gender		Yes	8	8
State/Region × Age × Gender		Yes		
State/Region × Age × Hispanicity		Yes		
State/Region × Age × Race (3 Levels)		Yes		
State/Region × Hispanicity × Gender		Yes		
State/Region × Race (3 Levels) × Hispanicity		Yes		
State/Region × Race (3 Levels) × Gender		Yes		

Appendix C.1: Model Group 1: Northeast

(Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont)

Table C.1a 2012 QDU Weight GEM Modeling Summary (Model Group 1: Northeast)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	2.33	3.79	1.09	2.9132	243	(0.59, 2.20)	(0.61, 2.20)
	2.15	4.48	1.06	2.9101	242	(0.34, 4.19)	(0.36, 4.19)
						(0.90, 3.03)	(0.90, 3.03)
res.qdu.nr	2.33	4.92	1.27	3.0250	243	(1.00, 2.50)	(1.00, 2.50)
	1.92	4.56	0.86	3.3407	236	(1.00, 4.61)	(1.00, 4.58)
						(1.00, 5.00)	(1.00, 1.17)
res.qdu.ps	1.92	4.56	0.86	3.3407	243	(0.81, 2.90)	(0.82, 2.90)
	2.06	4.21	0.68	3.3402	239	(0.58, 2.90)	(0.60, 2.70)
						(0.90, 1.03)	(0.90, 1.02)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13	duwght14	1-14
Minimum	18	1.00	18	0.29	8	0.38	8	0.54	6
1%	80	1.00	89	0.58	89	1.00	98	0.73	96
5%	120	1.00	162	0.74	164	1.01	179	0.93	173
10%	189	1.00	233	0.81	229	1.04	260	0.96	258
25%	305	1.00	503	0.91	483	1.12	521	0.99	513
Median	603	1.17	856	0.98	850	1.23	988	1.00	985
75%	834	3.71	1,954	1.07	1,931	1.37	2,287	1.01	2,308
90%	1,226	6.88	4,304	1.20	4,249	1.54	5,546	1.03	5,531
95%	1,576	7.89	6,266	1.31	6,144	1.65	8,446	1.05	8,442
99%	2,573	12.47	11,326	1.72	12,088	2.10	17,320	1.13	17,225
Maximum	7,128	15.74	30,862	7.32	29,264	4.58	47,848	2.86	45,340
n	12,616	-	12,616	-	12,616	-	9,917	-	9,917
Mean	679	2.62	1,718	1.01	1,705	1.27	2,168	1.00	2,168
Max/Mean	11	-	18	7.27	17	3.59	22	2.86	21

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 1 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 243 proposed effects, 242 were kept in the model, with the exception of State by Race, which combined American Indian or Alaska Native and Asian for New Hampshire.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

Out of 243 proposed effects, 236 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for each of the following States: New Jersey, Connecticut, New York, and Rhode Island. Three-factor effects for State/Region by Race by Hispanicity combined Black or African American with Other.

Respondent Questionnaire Dwelling Unit-Level Poststratification

Out of 243 proposed effects, 239 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for each of the following States: New Jersey, Connecticut, Vermont, and New Hampshire. All main and three-factor effects were retained at proposed levels.

Exhibit C.1.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	8	8	All levels present.
State (Binary)	9	8	8	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**132**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	31	Coll. (4,3) & (4,4); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**242**	

Exhibit C.1.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	8	8	All levels present.
State (Binary)	9	8	8	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**127**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	26	Coll. (6,3) & (6,4), (1,3) & (1,4), (5,3) & (5,4) & (5,5), (8,3) & (8,4) & (8,5); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**49**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	3	Coll. (1,2,1) & (1,3,1); conv.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**236**	

Exhibit C.1.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 1: Northeast

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**60**	**60**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	8	8	All levels present.
State (Binary)	9	8	8	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**133**	**129**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	9 × 5	32	32	All levels present.
State × Race	9 × 5	32	28	Coll. (1,3) & (1,4), (4,3) & (4,4), (5,3) & (5,4), (9,3) & (9,4); conv.
State × Gender	9 × 2	8	8	All levels present.
State × Hispanicity	9 × 2	8	8	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	3	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**243**	**239**	

Appendix C.2: Model Group 2: Midwest

(Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin)

Table C.2a 2012 QDU Weight GEM Modeling Summary (Model Group 2: Midwest)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	1.78	1.86	0.42	2.6793	300	(0.42, 1.10)	(0.42, 1.10)
	1.10	0.98	0.16	2.7109	300	(0.20, 3.27)	(0.20, 3.27)
						(0.90, 2.07)	(0.90, 2.07)
res.qdu.nr	1.25	1.26	0.20	2.7756	300	(1.00, 1.30)	(1.00, 1.30)
	0.76	0.90	0.12	2.9433	299	(1.00, 3.43)	(1.00, 3.33)
						(1.10, 1.56)	(1.10, 1.55)
res.qdu.ps	0.76	0.90	0.12	2.9433	300	(0.91, 1.20)	(0.92, 1.20)
	0.81	0.97	0.05	2.9433	300	(0.77, 1.21)	(0.78, 1.21)
						(0.90, 1.03)	(0.90, 1.03)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13	duwght14	1-14
Minimum	33	1.00	40	0.19	23	0.43	30	0.56	27
1%	97	1.00	107	0.61	109	1.00	128	0.96	128
5%	147	1.00	217	0.78	210	1.02	242	0.98	241
10%	268	1.00	373	0.85	350	1.05	385	0.99	382
25%	452	1.00	513	0.94	507	1.13	600	1.00	601
Median	541	1.13	737	1.00	762	1.21	911	1.00	913
75%	707	3.64	1,773	1.07	1,751	1.32	2,041	1.00	2,047
90%	1,150	6.65	3,754	1.16	3,805	1.44	4,917	1.01	4,913
95%	1,385	7.56	5,434	1.25	5,460	1.53	6,977	1.02	6,977
99%	1,832	11.84	9,852	1.59	10,371	1.81	14,335	1.05	14,315
Maximum	6,654	13.41	43,516	3.56	25,191	3.33	35,530	1.29	35,530
n	16,984	-	16,984	-	16,984	-	13,687	-	13,687
Mean	633	2.50	1,544	1.01	1,558	1.24	1,933	1.00	1,933
Max/Mean	11	-	28	-	16	-	18	-	18

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 2 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

All 300 proposed effects were kept in the model.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All main and three-factor effects were maintained at proposed levels. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for North Dakota.

Respondent Questionnaire Dwelling Unit-Level Poststratification

All 300 proposed effects were kept in the model.

Exhibit C.2.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	11	11	All levels present.
State (Binary)	9	11	11	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**163**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	44	All levels present.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity or Latino × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity or Latino × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**300**	

Exhibit C.2.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	11	11	All levels present.
State (Binary)	9	11	11	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**162**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	43	Coll. (9,3) & (9,4); conv.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	3	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**299**	

Exhibit C.2.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 2: Midwest

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**66**	**66**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	9	11	11	All levels present.
State (Binary)	9	11	11	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**163**	**163**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	12 × 5	44	44	All levels present.
State × Race	12 × 5	44	44	All levels present.
State × Gender	12 × 2	11	11	All levels present.
State × Hispanicity	12 × 2	11	11	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**71**	**71**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Hispanicity	4 × 5 × 2	12	12	All levels present.
State/Region × Age × Race (3 Levels)	4 × 5 × 3	24	24	All levels present.
State/Region × Hispanicity × Gender	4 × 2 × 2	3	3	All levels present.
State/Region × Race (3 Levels) × Hispanicity	4 × 3 × 2	6	6	All levels present.
State/Region × Race (3 Levels) × Gender	4 × 3 × 2	6	6	All levels present.
Total		**300**	**300**	

Appendix C.3: Model Group 3: South

(Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, West Virginia)

Table C.3a 2012 QDU Weight GEM Modeling Summary (Model Group 3: South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	1.50	1.91	0.38	2.5352	339	(0.42, 1.10)	(0.42, 1.10)
	1.17	1.51	0.17	2.5423	338	(0.46, 4.94)	(0.49, 4.93)
						(0.90, 4.99)	(0.90, 4.99)
res.qdu.nr	1.24	1.57	0.20	2.5925	339	(1.00, 1.70)	(1.00, 1.70)
	1.10	1.49	0.24	2.8943	339	(1.00, 5.00)	(1.00, 5.00)
						(1.30, 4.97)	(1.30, 4.79)
res.qdu.ps	1.10	1.49	0.24	2.8943	339	(0.20, 1.50)	(0.94, 1.50)
	1.09	1.51	0.12	2.8950	338	(0.20, 5.00)	(0.79, 2.79)
						(0..90, 5.00)	(0.95, 1.07)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C.3b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 3: South)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13	duwght14	1-14
Minimum	9	1.00	12	0.25	10	0.52	20	0.49	25
1%	59	1.00	75	0.65	71	1.00	75	0.94	75
5%	112	1.00	216	0.80	213	1.02	219	0.98	220
10%	253	1.00	425	0.85	412	1.05	449	0.99	444
25%	641	1.00	817	0.92	813	1.10	911	1.00	912
Median	923	1.15	1,284	1.00	1,300	1.18	1,479	1.00	1,482
75%	1,313	3.50	2,761	1.09	2,787	1.29	2,980	1.00	2,975
90%	1,716	6.54	6,014	1.19	6,214	1.43	7,732	1.01	7,712
95%	2,144	7.92	8,523	1.27	8,924	1.53	11,901	1.02	11,911
99%	3,118	12.02	14,217	1.59	14,457	1.82	19,919	1.06	19,913
Maximum	8,236	21.12	45,296	5.31	44,158	5.00	64,497	2.79	63,870
n	18,345	-	18,345	-	18,345	-	15,019	-	15,019
Mean	1,013	2.55	2,405	1.02	2,444	1.22	2,985	1.00	2,985
Max/Mean	8	-	19	-	18	-	22	-	21

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 3 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 339 proposed effects, 338 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for Delaware.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

All 339 proposed effects were kept in the model.

Respondent Questionnaire Dwelling Unit-Level Poststratification

Out of 339 proposed effects, 338 were kept in the model. Two-factor effects were modified for State by Race, combining American Indian or Alaska Native and Asian for Delaware.

Exhibit C.3.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	17	16	16	All levels present.
State (Binary)	17	16	16	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**212**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	63	Coll. (3,4) & (3,5); conv.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 Levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**338**	

Exhibit C.3.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	17	16	16	All levels present.
State (Binary)	17	16	16	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**213**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	64	All levels present.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**339**	

Exhibit C.3.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 3: South

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	17	16	16	All levels present.
State (Binary)	17	16	16	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**213**	**212**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	17 × 5	64	64	All levels present.
State × Race	17 × 5	64	63	Coll. (3,3) & (3,4); conv.
State × Gender	17 × 2	16	16	All levels present.
State × Hispanicity	17 × 2	16	16	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	8	All levels present.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**50**	**50**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Hispanicity	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Race (3 Levels)	3 × 5 × 3	16	16	All levels present.
State/Region × Hispanicity × Gender	3 × 2 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Hispanicity	3 × 3 × 2	4	4	All levels present.
State/Region × Race (3 levels) × Gender	3 × 3 × 2	4	4	All levels present.
Total		**339**	**338**	

Appendix C.4: Model Group 4: West

(Alaska, Arizona, California, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming)

Table C.4a 2012 QDU Weight GEM Modeling Summary (Model Group 4: West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Outwinsor			Nominal	Realized
sel.qdu.ps	2.09	3.52	0.99	3.1857	270	(0.49, 2.50)	(0.49, 2.50)
	1.44	2.60	0.62	3.1282	268	(0.34, 3.27)	(0.34, 3.27)
						(0.90, 1.66)	(0.90, 1.66)
res.qdu.nr	1.51	3.11	0.72	3.1508	270	(1.00, 2.90)	(1.00, 2.90)
	1.10	3.36	0.66	3.5391	267	(1.00, 5.00)	(1.00, 5.00)
						(1.30, 1.34)	(1.30, 1.34)
res.qdu.ps	1.10	3.36	0.66	3.5391	270	(0.82, 2.70)	(0.84, 2.70)
	0.91	2.98	0.57	3.5469	268	(0.71, 2.70)	(0.75, 1.71)
						(0.90, 1.02)	(1.02, 1.02)

GEM = generalized exponential model; QDU = questionnaire dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] There are six sets of bounds for each modeling step. Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The set of three bounds listed for each step correspond to the high extreme values, the nonextreme values, and the low extreme values.

Table C.4b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 4: West)

	SDU Weight	QDU Design Weight		sel.qdu.ps[1]		res.qdu.nr[1]		res.qdu.ps[1]	
	1-10	duwght11	1-11	duwght12	1-12	duwght13	1-13	duwght14	1-14
Minimum	20	1.00	20	0.22	13	0.45	41	0.38	44
1%	87	1.00	96	0.61	92	1.00	98	0.87	101
5%	112	1.00	141	0.74	132	1.02	155	0.96	155
10%	142	1.00	182	0.79	178	1.05	208	0.98	209
25%	251	1.00	426	0.88	416	1.11	468	0.99	469
Median	707	1.16	1,103	0.98	1,083	1.20	1,228	1.00	1,222
75%	1,480	3.32	2,176	1.09	2,200	1.32	2,645	1.01	2,627
90%	1,952	6.05	5,347	1.21	5,195	1.47	6,245	1.02	6,246
95%	2,211	7.90	8,100	1.31	8,010	1.57	10,398	1.04	10,482
99%	2,994	12.50	15,867	1.62	15,947	1.92	21,382	1.18	21,060
Maximum	8,444	14.78	48,558	3.77	35,038	5.00	53,491	1.71	59,335
n	12,676	-	12,676	-	12,676	-	10,227	-	10,227
Mean	918	2.43	2,109	1.00	2,080	1.24	2,578	1.00	2,578
Max/Mean	9	-	23	-	17	-	21	-	23

QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 4 Overview

Selected Questionnaire Dwelling Unit-Level Poststratification

Out of 270 proposed effects, 268 were kept in the model. All main effects were maintained in full. Two-factor effects were modified for Rent/Housing by Percent Black or African American, combining 50-100% and 10-<50% for the first and the fourth quintiles.

Respondent Questionnaire Dwelling Unit-Level Nonresponse

Out of 270 proposed effects, 267 were kept in the model. All main effects were maintained in full. Two-factor effects were modified for Rent/Housing by Percent Black or African American, combining 50-100% and 10-<50% for the first and the fourth quintiles. Three-factor effects were modified for State by Race by Hispanicity, combining Black or African American with Other for California.

Respondent Questionnaire Dwelling Unit-Level Poststratification

This step used the same set of effects as the selected questionnaire dwelling unit-level poststratification.

Exhibit C.4.1 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (sel.qdu.ps) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	13	12	12	All levels present.
State (Binary)	13	12	12	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**171**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	13 × 5	48	48	All levels present.
State × Race	13 × 5	48	48	All levels present.
State × Gender	13 × 2	12	12	All levels present.
State × Hispanicity	13 × 2	12	12	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	6	Coll. (1,1) & (2,1); zero. Coll. (1,4) & (2,4); sing.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**29**	**29**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Hispanicity	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Race (3 Levels)	2 × 5 × 3	8	8	All levels present.
State/Region × Hispanicity × Gender	2 × 2 × 2	1	1	All levels present.
State/Region × Race (3 Levels) × Hispanicity	2 × 3 × 2	2	2	All levels present.
State/Region × Race (3 Levels) × Gender	2 × 3 × 2	2	2	All levels present.
Total		**270**	**268**	

Exhibit C.4.2 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.nr) Model Group 4: West

Variables	Levels	Proposed	Final	Comments
One-Factor Effects		**68**	**68**	
Intercept	1	1	1	All levels present.
Group Quarter	3	2	2	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Household Type	7	6	6	All levels present.
Household Size	1	1	1	All levels present.
Rent/Housing	5	4	4	All levels present.
Population Density	4	3	3	All levels present.
% Black or African American	3	2	2	All levels present.
% Hispanic or Latino	35	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
State (Count)	13	12	12	All levels present.
State (Binary)	13	12	12	All levels present.
Quarter (Count)	4	3	3	All levels present.
Quarter (Binary)	4	3	3	All levels present.
Age Group	5	4	4	All levels present.
Race	5	4	4	All levels present.
Hispanicity	2	1	1	All levels present.
Gender	2	1	1	All levels present.
Two-Factor Effects		**173**	**171**	
Age × Race (3 Levels)	5 × 3	8	8	All levels present.
Age × Hispanicity	5 × 2	4	4	All levels present.
Age × Gender	5 × 2	4	4	All levels present.
Race (3 Levels) × Hispanicity	3 × 2	2	2	All levels present.
Race (3 Levels) × Gender	3 × 2	2	2	All levels present.
Hispanicity × Gender	2 × 2	1	1	All levels present.
State × Age	13 × 5	48	48	All levels present.
State × Race	13 × 5	48	48	All levels present.
State × Gender	13 × 2	12	12	All levels present.
State × Hispanicity	13 × 2	12	12	All levels present.
% Black or African American × % Owner-Occupied	3 × 3	4	4	All levels present.
% Black or African American × Rent/Housing	3 × 5	8	6	Coll. (1,1) & (2,1); zero. Coll. (1,4) & (2,4); sing.
% Hispanicity × % Owner-Occupied	3 × 3	4	4	All levels present.
% Hispanicity × Rent/Housing	3 × 5	8	8	All levels present.
% Owner-Occupied × Rent/Housing	3 × 5	8	8	All levels present.
Three-Factor Effects		**29**	**28**	
Race (3 Levels) × Age × Gender	3 × 5 × 2	8	8	All levels present.
State/Region × Age × Gender	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Hispanicity	2 × 5 × 2	4	4	All levels present.
State/Region × Age × Race (3 Levels)	2 × 5 × 3	8	8	All levels present.
State/Region × Hispanicity × Gender	2 × 2 × 2	1	1	All levels present.
State/Region × Race (3 Levels) × Hispanicity	2 × 3 × 2	2	1	Coll. (3,2,1) & (3,3,1); conv.
State/Region × Race (3 Levels) × Gender	2 × 3 × 2	2	2	All levels present.
Total		**270**	**267**	

Exhibit C.4.3 Covariates for 2012 NSDUH Questionnaire Dwelling Unit Weights (res.qdu.ps) Model Group 4: West

This step used the same set of covariates as the selected questionnaire dwelling unit poststratification.

Appendix D: Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Response Rates

Table D.1 2012 NSDUH QDU-Level Response Rates

Domain	Selected QDU	Respondent QDU	% Interview Response Rate[1]
Total	60,621	48,850	75.87
Census Region			
Northeast	12,616	9,917	73.10
South	18,345	15,019	76.39
Midwest	16,984	13,687	77.25
West	12,676	10,227	75.87
Quarter			
Quarter 1	14,387	11,628	75.80
Quarter 2	16,164	12,975	76.14
Quarter 3	15,762	12,743	76.00
Quarter 4	14,308	11,504	75.55
Household Type			
12-17, 18-25, 26+	5,811	5,004	85.32
12-17, 18-25	88	72	78.98
12-17, 26+	17,350	14,625	84.38
18-25, 26+	12,807	10,290	79.79
12-17	29	24	81.94
18-25	6,367	5,447	85.41
26+	18,169	13,388	72.47
Race/Ethnicity of Householder			
Hispanic or Latino White	7,920	6,574	78.41
Hispanic or Latino Black or African American	187	165	78.73
Hispanic or Latino Other	415	355	83.80
Non-Hispanic or Latino White	40,460	32,098	74.78
Non-Hispanic or Latino Black or African American	7,135	6,092	81.01
Non-Hispanic or Latino Other	4,504	3,566	72.15
% Hispanic or Latino in Segment			
50-100%	4,137	3,397	78.21
10-<50%	11,129	9,003	75.80
<10%	45,355	36,450	75.66
% Black or African American in Segment			
50-100%	4,381	3,749	80.55
10-<50%	9,185	7,524	76.96
<10%	47,055	37,577	75.17
% Owner-Occupied DUs in Segment			
50-100%	45,901	36,704	75.09
10-<50%	11,429	9,421	77.84
<10%	3,291	2,725	79.98
Combined Median Rent/Housing Value			
1st Quintile	9,920	8,288	78.90
2nd Quintile	12,960	10,661	77.97
3rd Quintile	13,713	11,009	75.45
4th Quintile	13,171	10,428	74.99
5th Quintile	10,857	8,464	73.22
Population Density			
Large MSA	25,508	20,223	74.24
Medium to Small MSA	30,307	24,705	77.53
Non-MSA, Urban	1,249	1,028	80.00
Non-MSA, Rural	3,557	2,894	77.06
Group Quarters			
Group	774	729	93.20
Non-Group	59,847	48,121	75.76
Household Size			
One	7,063	5,567	75.11
Two	21,700	16,974	73.66
Three	17,330	14,071	78.45
Four or More	14,528	12,238	82.73

DU = dwelling unit; MSA = metropolitan statistical area; QDU = questionnaire dwelling unit; SDU = screener dwelling unit.

[1] The weight used for calculating the response rate includes SDU- and QDU-level design weights, SDU nonresponse and poststratification adjustments, and selected QDU poststratification adjustment. This weight is the product of WT1*...*WT9*DUWT10*DUWT11.

Appendix E: Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Proportions of Extreme Values and Outwinsors

Table E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.qdu.ps[1] (SDUWT*DUWT10)			After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	60,621	2.62	5.38	1.44	1.88	2.60	0.66	1.41	2.17	0.43
Census Region										
Northeast	12,616	2.85	6.54	2.01	2.33	3.79	1.09	2.15	4.48	1.06
South	18,345	2.15	4.78	1.13	1.50	1.91	0.38	1.17	1.51	0.17
Midwest	16,984	2.47	4.31	0.91	1.78	1.86	0.42	1.10	0.98	0.16
West	12,676	3.26	6.46	1.98	2.09	3.52	0.99	1.44	2.60	0.62
Quarter										
Quarter 1	14,387	2.91	5.41	1.48	1.97	2.64	0.70	1.46	2.27	0.50
Quarter 2	16,164	2.36	5.15	1.41	1.68	2.28	0.59	1.39	2.35	0.49
Quarter 3	15,762	2.33	4.78	1.12	1.76	2.49	0.54	1.21	1.95	0.33
Quarter 4	14,308	2.93	6.19	1.75	2.12	3.00	0.81	1.60	2.10	0.39
Household Type										
12-17, 18-25, 26+	5,811	2.63	5.43	1.52	2.63	5.43	1.52	2.27	5.20	1.09
12-17, 18-25	88	4.55	7.61	1.34	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	17,350	2.66	5.20	1.42	2.64	5.18	1.42	1.98	4.21	0.89
18-25, 26+	12,807	2.96	6.18	1.63	2.46	5.57	1.47	1.56	4.39	1.12
12-17	29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18-25	6,367	2.42	4.70	0.96	2.18	4.44	0.91	1.66	3.42	0.61
26+	18,169	2.39	5.12	1.43	0.40	1.35	0.32	0.40	1.16	0.18
Race/Ethnicity of Householder										
Hispanic or Latino White	7,920	2.45	5.03	1.41	1.93	2.74	0.80	1.40	2.28	0.41
Hispanic or Latino Black or African American	187	57.75	82.43	35.31	48.13	60.13	21.66	40.64	61.93	18.69
Hispanic or Latino Other	415	34.70	62.52	23.51	27.95	44.43	11.95	17.83	31.05	6.77
Non-Hispanic or Latino White	40,460	1.26	2.31	0.43	0.80	1.01	0.18	0.49	0.68	0.08
Non-Hispanic or Latino Black or African American	7,135	3.90	7.48	1.82	3.10	4.13	1.04	3.27	4.20	0.80
Non-Hispanic or Latino Other	4,504	7.88	11.84	2.62	5.20	5.19	1.23	3.60	4.59	0.96

Table E.1 2012 NSDUH Selected QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.qdu.ps[1] (SDUWT*DUWT10)			After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
% Hispanic or Latino in Segment										
50-100%	4,137	2.13	5.12	1.88	1.84	3.56	1.13	1.55	3.46	1.00
10-<50%	11,129	3.80	8.36	2.68	2.95	4.25	1.26	2.40	3.64	0.81
<10%	45,355	2.37	4.37	0.94	1.62	2.01	0.43	1.15	1.59	0.25
% Black or African American in Segment										
50-100%	4,381	3.54	8.02	2.26	3.15	4.93	1.21	3.99	6.13	1.37
10-<50%	9,185	2.77	6.22	1.97	2.19	3.75	1.15	2.01	3.58	0.79
<10%	47,055	2.50	4.89	1.22	1.70	2.13	0.50	1.05	1.47	0.26
% Owner-Occupied DUs in Segment										
50-100%	45,901	2.27	4.70	1.25	1.59	2.10	0.52	1.06	1.55	0.27
10-<50%	11,429	3.73	7.01	1.84	2.80	3.95	1.00	1.95	3.05	0.74
<10%	3,291	3.65	8.34	2.33	2.61	4.85	1.40	4.35	7.78	1.53
Combined Median Rent/Housing Value										
1st Quintile	9,920	2.46	4.70	1.26	1.88	2.29	0.60	1.76	2.24	0.42
2nd Quintile	12,960	2.35	5.03	1.55	1.87	2.70	0.66	1.40	1.94	0.47
3rd Quintile	13,713	2.68	5.38	1.47	1.90	2.65	0.65	1.24	2.07	0.39
4th Quintile	13,171	2.73	5.46	1.50	1.90	2.55	0.70	1.31	2.30	0.50
5th Quintile	10,857	2.88	6.09	1.33	1.83	2.72	0.66	1.43	2.28	0.35
Population Density										
Large MSA[1]	25,508	2.93	6.67	1.86	2.16	3.43	0.90	1.81	2.91	0.62
Medium to Small MSA[1]	30,307	2.47	4.22	1.06	1.75	1.84	0.46	1.17	1.53	0.24
Non-MSA,[1] Urban	1,249	1.28	1.23	0.16	0.96	0.41	0.06	0.88	0.54	0.09
Non-MSA,[1] Rural	3,557	2.08	2.19	0.38	1.27	1.37	0.12	0.79	0.54	0.12
Group Quarters										
Group	774	4.01	7.46	0.51	2.71	3.60	0.36	2.97	6.11	1.34
Non-Group	59,847	2.60	5.35	1.45	1.87	2.60	0.66	1.39	2.14	0.42
Household Size										
One	7,063	2.10	5.10	1.33	0.93	1.17	0.28	0.72	1.02	0.14
Two	21,700	2.35	5.09	1.38	1.35	2.17	0.51	1.01	1.70	0.32
Three	17,330	2.80	5.40	1.50	2.44	4.60	1.28	1.51	3.49	0.80
Four or More	14,528	3.06	5.85	1.47	2.46	4.72	1.21	2.21	4.72	1.00

[1] DU = dwelling unit, MSA = metropolitan statistical area, PS = poststratification adjustment, QDU = questionnaire dwelling unit, SDU = screener dwelling unit, Sel = selected.
[2] Weighted extreme value proportion: $100*\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values, and w_k denotes the weight for both extreme values and nonextreme values.
[3] Outwinsor weight proportion: $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table E.2 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)			After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	48,850	1.52	2.42	0.50	1.17	2.32	0.42	1.17	2.20	0.31
Census Region										
Northeast	9,917	2.33	4.92	1.27	1.92	4.56	0.86	2.06	4.21	0.68
South	15,019	1.24	1.57	0.20	1.10	1.49	0.24	1.09	1.51	0.12
Midwest	13,687	1.25	1.26	0.20	0.76	0.90	0.12	0.81	0.97	0.05
West	10,227	1.51	3.11	0.72	1.10	3.36	0.66	0.91	2.98	0.57
Quarter										
Quarter 1	11,628	1.49	2.55	0.56	1.18	2.58	0.43	1.22	2.49	0.32
Quarter 2	12,975	1.53	2.81	0.58	1.21	2.43	0.51	1.16	2.24	0.40
Quarter 3	12,743	1.37	2.08	0.38	0.94	2.01	0.33	0.97	1.91	0.22
Quarter 4	11,504	1.70	2.24	0.47	1.36	2.28	0.40	1.36	2.16	0.29
Household Type										
12-17, 18-25, 26+	5,004	2.16	5.05	1.03	1.60	4.54	1.05	1.58	4.10	0.75
12-17, 18-25	72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
12-17, 26+	14,625	2.18	4.63	0.99	1.42	3.71	0.86	1.41	3.42	0.67
18-25, 26+	10,290	1.73	4.53	1.18	1.62	4.91	1.17	1.66	4.67	0.91
12-17	24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
18-25	5,447	1.43	3.37	0.69	1.19	2.40	0.34	1.10	2.14	0.30
26+	13,388	0.44	1.31	0.22	0.39	1.50	0.18	0.41	1.46	0.11
Race/Ethnicity of Householder										
Hispanic or Latino White	6,574	1.55	2.76	0.40	1.11	2.30	0.43	1.22	2.36	0.34
Hispanic or Latino Black or African American	165	40.61	65.50	19.97	38.18	57.52	16.88	40.00	55.69	15.87
Hispanic or Latino Other	355	16.90	30.32	7.01	10.14	22.63	4.43	9.30	16.59	2.82
Non-Hispanic or Latino White	32,098	0.53	0.66	0.09	0.50	0.81	0.07	0.40	0.68	0.02
Non-Hispanic or Latino Black or African American	6,092	3.33	4.69	0.95	1.63	3.26	0.63	1.89	3.32	0.48
Non-Hispanic or Latino Other	3,566	3.95	5.81	1.33	3.93	8.90	1.64	4.21	9.25	1.14
% Hispanic or Latino in Segment										
50-100%	3,397	1.65	3.36	0.97	1.32	2.70	0.77	1.47	2.70	0.65
10-<50%	9,003	2.63	4.27	0.95	1.79	3.98	0.82	1.88	3.77	0.61
<10%	36,450	1.23	1.76	0.31	1.00	1.78	0.26	0.97	1.68	0.18

Table E.2 2012 NSDUH Respondent QDU-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)			After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)			Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
% Black or African American in Segment										
50-100%	3,749	4.27	7.07	1.64	2.37	6.14	1.03	2.69	5.77	0.77
10-<50%	7,524	2.21	3.91	0.87	1.66	3.50	0.80	1.70	3.25	0.65
<10%	37,577	1.11	1.60	0.30	0.95	1.69	0.27	0.91	1.62	0.18
% Owner-Occupied DUs in Segment										
50-100%	36,704	1.18	1.83	0.35	1.01	2.00	0.34	0.98	1.91	0.24
10-<50%	9,421	1.98	3.19	0.77	1.33	2.82	0.56	1.54	2.79	0.46
<10%	2,725	4.48	7.59	1.61	2.79	5.11	0.98	2.46	4.33	0.74
Combined Median Rent/Housing Value										
1st Quintile	8,288	1.91	2.67	0.44	0.98	1.47	0.32	0.92	1.37	0.25
2nd Quintile	10,661	1.52	2.19	0.52	0.92	1.77	0.43	1.01	1.76	0.42
3rd Quintile	11,009	1.38	2.24	0.48	1.15	2.40	0.38	1.13	2.09	0.20
4th Quintile	10,428	1.44	2.74	0.61	1.18	2.74	0.47	1.23	2.70	0.38
5th Quintile	8,464	1.42	2.30	0.41	1.68	2.90	0.45	1.59	2.76	0.25
Population Density										
Large MSA[1]	20,223	1.96	3.19	0.72	1.71	3.52	0.63	1.67	3.26	0.46
Medium to Small MSA[1]	24,705	1.27	1.83	0.31	0.81	1.03	0.20	0.86	1.08	0.15
Non-MSA,[1] Urban	1,028	0.78	0.53	0.10	0.58	0.33	0.05	0.49	0.32	0.05
Non-MSA,[1] Rural	2,894	0.83	0.60	0.16	0.66	1.80	0.20	0.59	1.64	0.07
Group Quarters										
Group	729	2.88	6.48	1.42	1.51	2.75	0.15	1.37	3.82	1.11
Non-Group	48,121	1.50	2.39	0.49	1.16	2.32	0.42	1.17	2.19	0.30
Household Size										
One	5,567	0.70	1.13	0.17	0.79	1.43	0.15	0.65	1.30	0.11
Two	16,974	1.10	2.01	0.42	0.95	2.00	0.31	0.98	1.89	0.21
Three	14,071	1.69	3.67	0.78	1.17	2.95	0.72	1.28	3.06	0.65
Four or More	12,238	2.27	4.85	1.07	1.64	4.72	1.04	1.54	4.26	0.65

[1] DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PS = poststratification adjustment, QDU = questionnaire dwelling unit, Res = Respondent, SDU = screener dwelling unit.

[2] Weighted extreme value proportion: $100*\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values, and w_k denotes the weight for both extreme values and nonextreme values.

[3] Outwinsor weight proportion: $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Appendix F: Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Slippage Rates

Table F.1 2012 NSDUH QDU-Level Slippage Rates

Domain	*n*	Initial Total (*I*)[1]	Final Total (*F*)[2]	Control from SDU Weights (*C*)	(*I* - *C*)/*C*%	(*F* - *C*)/*C*%
Total	48,850	119,147,983	119,147,983	119,147,983	0.00	-0.00
Census Region						
Northeast	9,917	21,504,846	21,504,846	21,504,846	0.00	-0.00
South	15,019	44,826,838	44,826,838	44,826,838	0.00	-0.00
Midwest	13,687	26,453,818	26,453,818	26,453,818	0.00	-0.00
West	10,227	26,362,481	26,362,481	26,362,481	0.00	-0.00
Quarter						
Quarter 1	11,628	29,829,432	29,829,432	29,829,432	0.00	-0.00
Quarter 2	12,975	29,634,723	29,634,723	29,634,723	0.00	-0.00
Quarter 3	12,743	29,744,017	29,744,017	29,744,017	0.00	-0.00
Quarter 4	11,504	29,939,811	29,939,811	29,939,811	0.00	-0.00
Household Type						
12-17, 18-25, 26+	5,004	5,062,357	5,062,357	5,062,357	0.00	0.00
12-17, 18-25	72	66,238	66,238	66,238	0.00	0.00
12-17, 26+	14,625	13,691,239	13,691,239	13,691,239	0.00	-0.00
18-25, 26+	10,290	13,956,293	13,956,293	13,956,293	0.00	-0.00
12-17	24	22,258	22,258	22,258	0.00	0.00
18-25	5,447	5,779,323	5,779,323	5,779,323	0.00	-0.00
26+	13,388	80,570,274	80,570,274	80,570,274	0.00	-0.00
Race/Ethnicity of Householder						
Hispanic or Latino White	6,574	13,508,511	13,508,511	13,508,511	-0.00	-0.00
Hispanic or Latino Black or African American	165	725,656	725,656	725,656	-0.00	-0.00
Hispanic or Latino Other	355	1,005,979	1,005,979	1,005,979	0.00	-0.00
Non-Hispanic or Latino White	32,098	81,939,462	81,939,462	81,939,462	0.00	-0.00
Non-Hispanic or Latino Black or African American	6,092	14,332,726	14,332,726	14,332,726	0.00	-0.00
Non-Hispanic or Latino Other	3,566	7,635,648	7,635,648	7,635,648	0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	3,397	8,634,013	8,634,013	8,634,013	0.00	-0.00
10-<50%	9,003	25,715,926	25,715,926	25,715,926	0.00	-0.00
<10%	36,450	84,798,044	84,798,044	84,798,044	0.00	-0.00
% Black or African American in Segment						
50-100%	3,749	8,810,957	8,810,957	8,810,957	0.00	-0.00
10-<50%	7,524	20,110,346	20,110,346	20,110,346	-0.00	-0.00
<10%	37,577	90,226,680	90,226,679	90,226,679	0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	36,704	90,409,242	90,409,242	90,409,242	0.00	-0.00
10-<50%	9,421	22,264,244	22,264,244	22,264,244	0.00	-0.00
<10%	2,725	6,474,497	6,474,497	6,474,497	0.00	-0.00

Table F.1 2012 NSDUH QDU-Level Slippage Rates (continued)

Domain	*n*	Initial Total (*I*)[1]	Final Total (*F*)[2]	Control from SDU Weights (*C*)	(*I* - *C*)/*C*%	(*F* - *C*)/*C*%
Combined Median Rent/Housing Value						
1st Quintile	8,288	17,328,372	17,328,372	17,328,372	0.00	-0.00
2nd Quintile	10,661	23,902,320	23,902,320	23,902,320	0.00	-0.00
3rd Quintile	11,009	24,983,991	24,983,991	24,983,991	0.00	-0.00
4th Quintile	10,428	27,292,895	27,292,895	27,292,895	0.00	-0.00
5th Quintile	8,464	25,640,405	25,640,405	25,640,405	0.00	-0.00
Population Density						
Large MSA	20,223	60,608,068	60,608,068	60,608,068	0.00	-0.00
Medium to Small MSA	24,705	50,803,819	50,803,818	50,803,818	0.00	-0.00
Non-MSA, Urban	1,028	2,013,232	2,013,232	2,013,232	0.00	-0.00
Non-MSA, Rural	2,894	5,722,865	5,722,865	5,722,865	0.00	-0.00
Group Quarters						
Group	729	766,775	766,775	766,775	0.00	-0.00
Non-Group	48,121	118,381,208	118,381,208	118,381,208	0.00	-0.00
Household Size						
One	5,567	31,354,646	31,358,657	31,327,228	0.09	0.10
Two	16,974	54,585,381	54,556,667	54,622,573	-0.07	-0.12
Three	14,071	19,065,921	19,111,984	19,137,226	-0.37	-0.13
Four or More	12,238	14,142,035	14,120,675	14,060,955	0.58	0.42

DU = dwelling unit, MSA = metropolitan statistical area, QDU = questionnaire dwelling unit, SDU = screener dwelling unit.

[1] WT1*...*WT9*DUWT10*...*DUWT12 (before QDU poststratification).

[2] WT1*...*WT9*DUWT10*...*DUWT13 (after QDU poststratification).

Appendix G: Evaluation of Calibration Weights: Questionnaire Dwelling Unit-Level Weight Summary Statistics

Table G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.qdu.ps[1] (SDUWT*DUWT10)						After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	60,621	9	436	664	1,098	8,444	1.56	12	552	1,013	2,115	48,558	2.86	8	542	1,015	2,137	44,158	2.85
Census Region																			
Northeast	12,616	18	305	603	834	7,128	1.57	18	503	856	1,954	30,862	2.91	8	483	850	1,931	29,264	2.91
South	18,345	9	641	923	1,313	8,236	1.39	12	817	1,284	2,761	45,296	2.54	10	813	1,300	2,787	44,158	2.54
Midwest	16,984	33	452	541	707	6,654	1.35	40	513	737	1,773	43,516	2.68	23	507	762	1,751	25,191	2.71
West	12,676	20	251	707	1,480	8,444	1.75	20	426	1,103	2,176	48,558	3.19	13	416	1,083	2,200	35,038	3.13
Quarter																			
Quarter 1	14,387	21	495	717	1,140	8,082	1.54	21	610	1,079	2,216	45,473	2.88	10	597	1,065	2,231	35,038	2.85
Quarter 2	16,164	18	402	601	1,045	8,236	1.56	18	505	951	1,961	30,862	2.80	8	498	955	1,980	27,818	2.84
Quarter 3	15,762	9	431	617	1,087	7,138	1.57	12	520	981	2,068	45,296	2.85	10	510	975	2,077	44,158	2.85
Quarter 4	14,308	16	471	699	1,111	8,444	1.57	16	590	1,035	2,274	48,558	2.88	13	585	1,062	2,280	30,041	2.84
Household Type																			
12-17, 18-25, 26+	5,811	18	463	679	1,156	8,444	1.59	18	463	679	1,156	8,446	1.59	13	454	680	1,164	9,750	1.59
12-17, 18-25	88	42	357	592	958	2,459	1.53	42	357	592	958	2,460	1.53	27	350	559	1,099	2,529	1.61
12-17, 26+	17,350	16	412	626	1,059	8,351	1.55	16	414	628	1,065	8,379	1.56	10	400	631	1,080	8,263	1.56
18-25, 26+	12,807	9	486	709	1,188	8,236	1.57	12	584	891	1,410	9,795	1.53	8	569	905	1,441	11,391	1.54
12-17	29	83	319	641	951	2,460	1.52	84	322	648	957	2,485	1.52	116	322	579	1,205	2,291	1.56
18-25	6,367	21	364	652	1,069	6,940	1.58	21	424	784	1,237	7,503	1.52	30	408	768	1,226	6,586	1.56
26+	18,169	20	435	656	1,080	8,082	1.53	74	1,893	3,509	5,773	48,558	1.72	53	1,865	3,491	5,815	44,158	1.71
Race/Ethnicity of Householder																			
Hispanic or Latino White	7,920	19	505	809	1,317	8,444	1.44	20	596	1,102	1,839	27,837	2.38	8	589	1,110	1,858	26,638	2.47
Hispanic or Latino Black or African American	187	18	807	1,559	2,869	8,351	1.79	18	947	2,014	4,659	37,254	2.79	20	896	1,913	4,342	29,264	2.76
Hispanic or Latino Other	415	9	189	615	1,448	7,653	2.68	12	269	935	2,525	45,473	4.61	10	298	964	2,404	33,561	4.15
Non-Hispanic or Latino White	40,460	18	416	625	1,016	8,082	1.50	18	537	992	2,214	48,558	2.86	10	530	985	2,217	44,158	2.87
Non-Hispanic or Latino Black or African American	7,135	29	549	757	1,191	7,579	1.50	29	667	1,046	2,111	43,516	2.80	13	656	1,078	2,186	31,554	2.69
Non-Hispanic or Latino Other	4,504	20	244	595	1,222	8,359	1.79	20	383	924	2,043	39,281	3.04	21	362	896	1,942	35,038	3.03
% Hispanic or Latino in Segment																			
50-100%	4,137	26	590	1,011	1,441	6,622	1.34	26	769	1,331	2,091	22,955	2.35	33	763	1,348	2,139	26,638	2.43
10-<50%	11,129	9	565	876	1,420	8,444	1.49	12	712	1,317	2,491	43,516	2.57	10	712	1,330	2,557	30,041	2.56
<10%	45,355	16	384	605	963	8,082	1.57	16	512	916	2,005	48,558	3.00	8	503	913	1,999	44,158	2.98

Table G.1 2012 NSDUH Selected QDU-Level Weight Summary Statistics (continued)

Domain	n	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.qdu.ps[1] (SDUWT*DUWT10)						After sel.qdu.ps[1] (SDUWT*DUWT10*DUWT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
% Black or African American in Segment																			
50-100%	4,381	9	530	735	1,115	7,874	1.57	12	648	1,002	2,038	27,138	2.99	10	650	1,050	2,184	31,554	2.87
10-<50%	9,185	22	533	802	1,286	8,236	1.48	22	688	1,192	2,412	43,516	2.65	13	680	1,197	2,432	30,807	2.59
<10%	47,055	18	400	627	1,055	8,444	1.57	18	520	975	2,070	48,558	2.89	8	511	970	2,071	44,158	2.91
% Owner-Occupied DUs[1] in Segment																			
50-100%	45,901	18	434	648	1,074	8,444	1.55	18	546	1,005	2,131	48,558	2.85	8	540	1,011	2,143	44,158	2.84
10-<50%	11,429	16	433	698	1,115	8,236	1.57	16	567	1,006	2,055	43,516	2.98	13	541	1,000	2,040	35,038	2.97
<10%	3,291	9	473	810	1,317	8,351	1.59	12	599	1,135	2,145	37,254	2.64	10	585	1,150	2,315	27,818	2.64
Combined Median Rent/Housing Value																			
1st Quintile	9,920	19	385	594	903	8,351	1.58	19	507	876	1,879	26,477	2.82	8	502	900	1,909	31,554	2.84
2nd Quintile	12,960	18	402	622	1,024	8,444	1.63	18	512	944	1,995	30,624	2.90	10	514	948	2,013	29,325	2.85
3rd Quintile	13,713	9	370	612	1,008	8,082	1.61	12	502	920	1,928	43,136	3.04	10	494	912	1,899	30,807	3.05
4th Quintile	13,171	16	474	732	1,182	7,853	1.53	16	609	1,110	2,291	45,296	2.81	21	585	1,083	2,279	44,158	2.81
5th Quintile	10,857	25	534	816	1,257	8,359	1.42	25	666	1,227	2,577	48,558	2.64	19	676	1,255	2,624	33,561	2.63
Population Density																			
Large MSA[1]	25,508	9	578	849	1,343	8,359	1.41	12	731	1,298	2,664	48,558	2.55	10	728	1,311	2,650	35,038	2.54
Medium to Small MSA[1]	30,307	18	315	558	886	8,444	1.62	18	458	838	1,765	43,516	3.06	8	450	843	1,770	30,041	3.06
Non-MSA,[1] Urban	1,249	20	207	512	791	3,628	1.63	20	382	770	1,684	16,479	2.97	45	359	792	1,711	17,924	3.05
Non-MSA,[1] Rural	3,557	18	199	469	802	4,220	1.71	19	335	772	1,692	45,296	3.44	23	330	759	1,648	44,158	3.51
Group Quarters																			
Group	774	25	219	459	986	4,098	1.90	25	275	592	1,269	30,862	4.92	27	264	571	1,225	20,208	3.38
Non-Group	59,847	9	439	665	1,099	8,444	1.56	12	556	1,019	2,136	48,558	2.85	8	548	1,021	2,153	44,158	2.84
Household Size																			
One	7,063	23	404	638	1,029	6,552	1.52	69	998	2,660	6,271	48,558	2.16	55	976	2,629	6,225	44,158	2.16
Two	21,700	19	428	649	1,055	8,236	1.54	20	706	1,490	3,530	45,473	2.21	14	704	1,481	3,501	33,561	2.20
Three	17,330	18	446	672	1,105	8,444	1.55	18	475	785	1,353	24,578	2.07	21	468	789	1,369	15,882	2.05
Four or More	14,528	9	442	694	1,185	8,359	1.60	12	454	729	1,268	14,597	1.78	8	439	728	1,262	13,876	1.79

[1] DU = dwelling unit, MSA = metropolitan statistical area, PS = poststratification adjustment, QDU = questionnaire dwelling unit, SDU = screener dwelling unit, Sel = selected.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Table G.2 2012 NSDUH Respondent QDU-Level Weight Summary Statistics

Domain	*n*	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)						After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	48,850	8	527	971	1,975	44,158	2.90	8	627	1,175	2,485	64,497	3.21	6	626	1,178	2,487	63,870	3.22
Census Region																			
Northeast	9,917	8	450	803	1,716	29,264	3.02	8	521	988	2,287	47,848	3.34	6	513	985	2,308	45,340	3.34
South	15,019	10	788	1,254	2,452	44,158	2.59	20	911	1,479	2,980	64,497	2.89	25	912	1,482	2,975	63,870	2.89
Midwest	13,687	23	502	739	1,634	25,191	2.78	30	600	911	2,041	35,530	2.94	27	601	913	2,047	35,530	2.94
West	10,227	41	397	1,024	2,093	35,038	3.15	41	468	1,228	2,645	53,491	3.54	44	469	1,222	2,627	59,335	3.55
Quarter																			
Quarter 1	11,628	10	581	1,024	2,051	35,038	2.93	10	693	1,249	2,586	53,491	3.22	8	693	1,251	2,566	59,335	3.22
Quarter 2	12,975	8	484	914	1,850	27,818	2.91	8	577	1,102	2,336	37,706	3.18	6	577	1,106	2,336	37,778	3.19
Quarter 3	12,743	10	498	930	1,935	44,158	2.86	15	587	1,106	2,410	64,497	3.20	13	588	1,108	2,415	63,870	3.20
Quarter 4	11,504	13	571	1,020	2,096	30,014	2.88	14	681	1,230	2,660	47,848	3.21	10	678	1,231	2,658	45,340	3.21
Household Type																			
12-17, 18-25, 26+	5,004	13	451	680	1,156	9,750	1.57	14	508	781	1,340	9,750	1.61	10	506	781	1,342	9,750	1.60
12-17, 18-25	72	27	301	536	1,099	2,529	1.69	27	329	701	1,375	3,411	1.68	22	321	697	1,349	3,856	1.74
12-17, 26+	14,625	10	398	632	1,081	8,263	1.57	10	458	747	1,266	10,398	1.60	8	457	748	1,267	12,788	1.61
18-25, 26+	10,290	8	558	903	1,430	9,930	1.54	8	673	1,112	1,778	15,051	1.60	6	673	1,115	1,775	20,359	1.62
12-17	24	116	279	611	1,208	1,725	1.48	124	295	836	1,446	2,127	1.46	121	301	847	1,435	2,135	1.47
18-25	5,447	30	403	761	1,221	6,586	1.57	33	455	892	1,434	6,277	1.57	34	454	889	1,432	12,046	1.59
26+	13,388	53	1,827	3,412	5,696	44,158	1.73	57	2,383	4,556	7,834	64,497	1.78	44	2,384	4,544	7,837	63,870	1.78
Race/Ethnicity of Householder																			
Hispanic or Latino White	6,574	8	576	1,077	1,743	26,638	2.49	8	648	1,251	2,105	35,972	2.86	6	648	1,254	2,098	36,038	2.88
Hispanic or Latino Black or African American	165	20	796	1,789	3,943	29,264	2.99	20	885	2,126	4,681	47,848	3.39	25	1,009	2,201	4,493	45,340	3.38
Hispanic or Latino Other	355	10	296	982	2,419	33,561	4.10	28	313	1,051	2,782	44,608	4.82	30	330	1,045	2,888	38,613	4.48
Non-Hispanic or Latino White	32,098	10	513	942	2,043	44,158	2.94	10	621	1,165	2,641	64,497	3.20	8	620	1,168	2,641	63,870	3.20
Non-Hispanic or Latino Black or African American	6,092	34	647	1,048	2,037	28,858	2.69	36	718	1,182	2,379	34,671	3.01	36	720	1,190	2,382	34,627	3.02
Non-Hispanic or Latino Other	3,566	23	344	833	1,782	35,038	3.05	27	408	1,040	2,373	53,491	3.63	22	397	1,028	2,369	59,335	3.69
% Hispanic or Latino in Segment																			
50-100%	3,397	33	731	1,300	2,006	26,638	2.48	33	845	1,510	2,440	35,972	2.78	35	850	1,502	2,433	36,038	2.79
10-<50%	9,003	10	689	1,277	2,377	29,264	2.58	20	801	1,528	3,041	47,848	2.89	23	801	1,536	3,041	45,340	2.89
<10%	36,450	8	488	880	1,851	44,158	3.04	8	580	1,069	2,333	64,497	3.35	6	580	1,069	2,333	63,870	3.35

Table G.2 2012 NSDUH Respondent QDU-Level Weight Summary Statistics (continued)

Domain	n	Before res.qdu.nr[1] (SDUWT*DUWT10*DUWT11)						After res.qdu.nr[1] (SDUWT*DUWT10*...*DUWT12)						Final Weight: After res.qdu.ps[1] (SDUWT*DUWT10*...*DUWT13)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
% Black or African American in Segment																			
50-100%	3,749	10	639	1,022	1,988	28,858	2.89	27	707	1,156	2,281	40,699	3.28	30	702	1,161	2,277	38,613	3.27
10-<50%	7,524	20	665	1,139	2,230	29,264	2.64	20	775	1,363	2,762	47,848	2.97	23	775	1,365	2,758	45,340	2.97
<10%	37,577	8	494	929	1,922	44,158	2.96	8	590	1,139	2,442	64,497	3.26	6	590	1,140	2,444	63,870	3.26
% Owner-Occupied DUs[1] in Segment																			
50-100%	36,704	8	526	967	1,972	44,158	2.89	8	632	1,180	2,510	64,497	3.19	6	631	1,181	2,510	63,870	3.19
10-<50%	9,421	23	517	954	1,924	35,038	3.04	26	594	1,130	2,309	53,491	3.38	17	594	1,131	2,312	59,335	3.40
<10%	2,725	10	574	1,109	2,223	27,818	2.67	20	646	1,292	2,720	39,554	2.93	25	634	1,293	2,718	37,754	2.93
Combined Median Rent/Housing Value																			
1st Quintile	8,288	8	487	871	1,784	28,858	2.86	8	563	1,014	2,149	35,972	3.16	6	564	1,010	2,139	36,038	3.17
2nd Quintile	10,661	10	499	915	1,897	26,430	2.92	10	582	1,085	2,290	40,699	3.24	8	582	1,086	2,287	38,613	3.24
3rd Quintile	11,009	10	481	883	1,765	29,155	3.06	20	569	1,069	2,237	39,581	3.41	20	568	1,072	2,242	39,310	3.41
4th Quintile	10,428	23	567	1,044	2,104	44,158	2.89	30	678	1,271	2,683	64,497	3.15	26	673	1,271	2,689	63,870	3.16
5th Quintile	8,464	19	646	1,194	2,378	33,561	2.67	20	792	1,491	3,123	44,608	2.93	18	789	1,491	3,109	40,087	2.93
Population Density																			
Large MSA[1]	20,223	10	704	1,243	2,431	35,038	2.60	20	843	1,499	3,145	53,491	2.88	17	843	1,502	3,128	59,335	2.88
Medium to Small MSA[1]	24,705	8	438	816	1,667	29,155	3.10	8	514	975	2,054	39,581	3.39	6	513	977	2,056	39,310	3.39
Non-MSA,[1] Urban	1,028	45	341	777	1,603	17,924	3.10	50	382	893	1,944	22,067	3.35	50	382	898	1,954	22,001	3.35
Non-MSA,[1] Rural	2,894	23	322	711	1,560	44,158	3.51	23	370	865	1,942	64,497	3.87	20	372	865	1,944	63,870	3.86
Group Quarters																			
Group	729	27	267	561	1,225	20,208	3.45	27	283	590	1,343	20,208	3.57	22	279	572	1,290	19,652	3.69
Non-Group	48,121	8	531	979	1,991	44,158	2.89	8	633	1,184	2,516	64,497	3.20	6	632	1,187	2,515	63,870	3.20
Household Size																			
One	5,567	63	917	2,379	5,997	44,158	2.22	72	1,114	2,949	8,054	64,497	2.34	83	1,114	2,965	8,041	63,870	2.35
Two	16,974	14	670	1,361	3,229	33,561	2.25	15	805	1,679	4,247	44,608	2.44	13	803	1,678	4,234	38,613	2.43
Three	14,071	23	459	775	1,341	15,882	2.01	27	543	938	1,617	23,531	2.32	22	544	940	1,621	24,616	2.34
Four or More	12,238	8	431	721	1,245	13,876	1.77	8	497	845	1,459	24,188	2.02	6	493	844	1,461	20,300	2.03

[1] DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PS = poststratification adjustment, QDU = questionnaire dwelling unit, Res = respondent, SDU = screener dwelling unit, Sel = selected.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Appendix H: GEM Modeling Summary for the Pair Weights

Appendix H: GEM Modeling Summary for the Pair Weights

This appendix summarizes each model group throughout all stages of weight calibration modeling. Unlike much of the other information presented in this report, this section provides a model-specific overview of weight calibration, as opposed to a domain-specific one.

For 2012, modeling involved taking two model groups through four adjustment steps: (1) selected pair poststratification, (2) pair nonresponse adjustment, (3) responding pair poststratification, and (4) responding pair extreme value adjustment.

Model-specific summary statistics are shown in Tables H.1a through H.2b. Included in these tables, for each stage of modeling, are the number of factor effects included in the final model; the high, low, and nonextreme weight bounds set to provide the upper and lower limits for the generalized exponential model (GEM) macro; the weighted, unweighted, and winsorized weight proportions; the unequal weighting effect (UWE); and weight distributions. The UWE provides an approximate partial measure of variance and provides a summary of how much impact a particular stage of modeling has on the distribution of the new product of weights. At each stage in the modeling, these summary statistics were calculated and utilized to help evaluate the quality of the weight component under the model chosen.

Occurrences of small sample sizes and exact linear combinations in the realized data led to situations whereby modeling inclusion of all originally proposed levels of covariates in the model was not possible. The text and exhibits in Sections H.1 and H.2 summarize the decisions made with regard to final covariates included in each model. For the list of proposed initial covariates considered at each stage of modeling, see Exhibit H.2. For the list of realized final model covariates, see Exhibits H.1.1 to H.2.4. For guidelines on interpreting these exhibits, see Appendix C.

Final Model Explanatory Variables

For brevity, numeric abbreviations for factor levels are established in Exhibit 4.2 (included here as Exhibit H.1 for easy reference). A complete list of all variables and associated levels used at any stage of modeling is provided. Note that not all factors or levels are present in all stages of modeling, and the initial set of variables is the same across model groups but may change for an adjustment step of modeling. The initial candidates are found in any of the proposed variables columns for a particular stage of weight adjustment.

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables

Group Quarter Indicator

1: College Dorm, 2: Other Group Quarter, 3: Non-Group Quarter[1]

Household Size

2: DU with 2 Persons,[1] 3: DU with 3 Persons, 4: DU with ≥ 4 Persons

Pair Age (15 Levels)

1: 12-17 and 12-17,[1] 2: 12-17 and 18-25, 3: 12-17 and 26-34, 4: 12-17 and 35-49, 5: 12-17 and 50+, 6: 18-25 and 18-25, 7: 18-25 and 26-34, 8: 18-25 and 35-49, 9: 18-25 and 50+, 10: 26-34 and 26-34, 11: 26-34 and 35-49, 12: 26-34 and 50+, 13: 35-49 and 35-49, 14: 35-49 and 50+, 15: 50+ and 50+

Pair Age (6 Levels)

1: 12-17 and 12-17,[1] 2: 12-17 and 18-25, 3: 12-17 and 26+, 4: 18-25 and 18-25, 5: 18-25 and 26+, 6: 26+ and 26+

Pair Age (3 Levels)

1: 12-17 and 12-17,[1] 2: 12-17 and 18+, 3: 18+ and 18+

Pair Gender

1: Male and Female,[1] 2: Female and Female, 3: Male and Male

Pair Race/Ethnicity (10 Levels)

1: White and White,[1] 2: White and Black or African American, 3: White and Hispanic or Latino, 4: White and Other, 5: Black or African American and Black or African American, 6: Black or African American and Hispanic or Latino, 7: Black or African American and Other, 8: Hispanic or Latino and Hispanic or Latino, 9: Hispanic or Latino and Other, 10: Other and Other

Pair Race/Ethnicity (5 Levels)

1: Two or More Races Pair, 2: Hispanic or Latino Pair, 3: Black or African American Pair, 4: White Pair,[1] 5: Other Pair

Pair Race/Ethnicity (4 Levels)

1: Two or More Races Pair or Other and Other, 2: Hispanic or Latino Pair, 3: Black or African American Pair, 4: White Pair[1]

Percentage of Owner-Occupied Dwelling Units in Segment (% Owner-Occupied)

1: 50-100%[1], 2: 10-<50%, 3: 0-<10%

Percentage of Segments That Are Black or African American

1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Percentage of Segments That Are Hispanic or Latino

1: 50-100%, 2: 10-<50%, 3: 0-<10%[1]

Segment-Combined Median Rent and Housing Value (Rent/Housing)[2]

1: First Quintile, 2: Second Quintile, 3: Third Quintile, 4: Fourth Quintile, 5: Fifth Quintile[1]

Population Density

1: MSA 1,000,000 or More, 2: MSA Less than 1,000,000, 3: Non-MSA Urban, 4: Non-MSA Rural[1]

Quarter

1: Quarter 1, 2: Quarter 2, 3: Quarter 3, 4: Quarter 4[1]

Race/Ethnicity of Householder

1: Hispanic or Latino White,[1] 2: Hispanic or Latino Black or African American, 3: Hispanic or Latino Other, 4: Non-Hispanic or Latino White, 5: Non-Hispanic or Latino Black or African American, 6: Non-Hispanic or Latino Other

State/Region

Model Group 1: 1: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, Rhode Island, Vermont; 2: Alabama, Arkansas, Delaware, District of Columbia, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Virginia, West Virginia;[1] 3: New York; 4: Pennsylvania; 5: Florida; 6: Texas

Model Group 2: 1: Indiana, Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, South Dakota, Wisconsin;[1] 2: Alaska, Arizona, Colorado, Idaho, Hawaii, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming; 3: Michigan; 4: Illinois; 5: Ohio; 6: California

Exhibit H.1 Definitions of Levels for Pair-Level Calibration Modeling Variables (continued)

States[3]

Model Group 1: 1: Alabama, 2: Arkansas, 3: Connecticut, 4: Delaware, 5: District of Columbia, 6: Florida, 7: Georgia, 8: Kentucky, 9: Louisiana, 10: Maine, 11: Maryland,[1] 12: Massachusetts, 13: Mississippi, 14: New Hampshire, 15: New Jersey, 16: New York, 17: North Carolina, 18: Oklahoma, 19: Pennsylvania, 20: Rhode Island, 21: South Carolina, 22: Tennessee, 23: Texas, 24: Vermont, 25: Virginia, 26: West Virginia

Model Group 2: 1: Alaska, 2: Arizona,[1] 3: California, 4: Colorado, 5: Idaho, 6: Illinois, 7: Indiana, 8: Iowa, 9: Hawaii, 10: Kansas, 11: Michigan, 12: Minnesota, 13: Missouri, 14: Montana, 15: Nebraska, 16: Nevada, 17: New Mexico, 18: North Dakota, 19: Ohio, 20: Oregon, 21: South Dakota, 22: Utah, 23: Washington, 24: Wisconsin, 25: Wyoming

Pair Relationship Associated with Multiplicity

1: Parent-Child (12-14)*
2: Parent-Child (12-17)*
3: Parent-Child (12-10)*
4: Parent*-Child (12-14)
5: Parent*-Child (12-17)
6: Parent*-Child (12-20)
7: Sibling (12-14)-Sibling (15-17)*
8: Sibling (12-17)-Sibling (18-25)*
9: Spouse-Spouse/Partner-Partner
10: Spouse-Spouse/Partner-Partner with Children (Younger than 18)

DU = dwelling unit, MSA = metropolitan statistical area.

[1] The reference level for this variable. This is the level against which effects of other factor levels are measured.

[2] Segment-Combined Median Rent and Housing Value is a composite measure based on rent, housing value, and percentage owner-occupied.

[3] The States or district assigned to a particular model is based on combined census regions.

* The pair member focused on.

Exhibit H.2 Covariates for 2012 NSDUH Pair Weights

Variables	Level	Proposed
One-Factor Effects		
Intercept	1	1
State	Model-specific	
Quarter	4	3
Population Density	3	2
Group Quarter	3	2
Household Size	3	2
Pair Age	15	14
Pair Gender	4	2
Pair Race/Ethnicity	10	9
Race/Ethnicity of Householder	6	5
Rent/Housing	5	4
Segment % Black or African American	3	2
Segment % Hispanic or Latino	3	2
% Owner-Occupied	3	2
Pair Relationship	Model-specific	
Two-Factor Effects		
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8
Pair Gender × Pair Age (6 Levels)	3 × 6	10
State/Region × Pair Race/Ethnicity (5 Levels)	Model-specific	
State/Region × Pair Age (6 Levels)	Model-specific	
State/Region × Pair Gender	Model-specific	
Rent/Housing × % Black or African American	5 × 3	8
Rent/Housing × % Hispanic or Latino	5 × 3	8
Rent/Housing × % Owner-Occupied	5 × 3	8
% Owner-Occupied × % Black or African American	3 × 3	4
% Owner-Occupied × % Hispanic or Latino	3 × 3	4
Three-Factor Effects		
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12

Appendix H.1: Model Group 1: Northeast and South

(Alabama, Arkansas, Connecticut, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maine, Massachusetts, Maryland, Mississippi, New Hampshire, New Jersey, New York, North Carolina, Oklahoma, Pennsylvania, Rhode Island, South Carolina, Tennessee, Texas, Vermont, Virginia, West Virginia)

Table H.1a 2012 Pair Weight GEM Modeling Summary (Model Group 1: Northeast and South)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Winsorized			Nominal	Realized
sel.pr.ps	5.38	23.39	12.45	67.5167	213	(0.26, 1.10)	(0.26, 1.10)
	1.86	7.18	1.41	8.5932	201	(0.21, 3.98)	(0.21, 3.96)
						(0.90, 1.60)	(0.90, 1.60)
res.pr.nr	1.74	6.26	1.31	8.9725	213	(1.00, 1.60)	(1.00, 1.60)
	2.09	9.12	1.85	9.3991	213	(1.00, 5.00)	(1.00, 5.00)
						N/A	N/A
res.pr.ps	2.09	9.39	1.81	9.3991	223	(0.39, 1.20)	(0.39, 1.20)
	1.47	5.44	0.55	9.6725	211	(0.25, 1.89)	(0.25, 1.89)
						N/A	N/A
res.pr.ev	1.47	5.44	0.55	9.6725	223	(0.94, 1.30)	(0.96, 1.30)
	0.81	5.70	0.19	9.6484	211	(0.80, 1.44)	(0.85, 1.28)
						N/A	N/A

GEM = generalized exponential model.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second is for nonextreme values, and the third is for low extreme values.

Table H.1b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 1: Northeast and South)

	SDU Weight	Pair Selection Prob		sel.pr.ps[1]		res.pr.nr[1]		res.pr.ps[1]		res.pr.ev[1]	
	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14	pairwt15	1-15
Minimum	9	1.02	20	0.01	8	0.41	10	0.17	4	0.61	3
1%	63	1.11	131	0.22	62	0.85	65	0.30	53	0.88	51
5%	127	1.20	355	0.31	216	1.00	229	0.49	189	0.95	184
10%	214	1.31	645	0.41	410	1.01	433	0.62	382	0.96	385
25%	498	1.51	1,307	0.63	1,075	1.04	1,145	0.84	1,091	0.98	1,085
Median	781	5.52	3,306	0.94	3,088	1.16	3,462	1.01	3,367	1.00	3,360
75%	1,185	11.85	8,863	1.30	8,660	1.42	10,165	1.15	9,887	1.01	9,890
90%	1,629	22.86	18,678	1.75	21,257	1.88	29,270	1.37	28,709	1.03	28,478
95%	2,096	28.63	28,096	2.13	33,961	2.33	51,092	1.52	51,237	1.04	51,790
99%	3,206	53.82	60,653	2.89	75,394	3.96	135,350	1.69	138,677	1.11	139,613
Maximum	7,138	4,403.70	6,068,764	3.96	1,053,229	5.00	1,323,156	1.89	1,234,015	1.28	1,237,337
n	13,619	-	13,619	-	13,619	-	9,723	-	9,723	-	9,723
Mean	905	10.01	8,700	1.03	8,841	1.34	12,384	1.00	12,384	1.00	12,384
Max/Mean	8	-	698	-	119	-	107	-	100	-	100

SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 1 Overview

Selected Pair-Level Poststratification

In the selected pair-level poststratification step, 201 of 213 proposed factors were retained in the final model. All main and two-factor effects were retained. All three-factor effects were dropped due to convergence problems.

Respondent Pair-Level Nonresponse

In the respondent pair-level nonresponse step, all proposed factors were retained in the final model.

Respondent Pair-Level Poststratification

In the respondent pair-level poststratification step, 211 of 223 proposed factors were retained in the final model. All main and two-factor effects were retained. All three-factor effects were dropped due to convergence problems.

Respondent Pair-Level Extreme Value Adjustment

This step used exactly the same variables as in the respondent pair-level nonresponse and poststratification steps.

Exhibit H.1.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**213**	**201**	

Exhibit H.1.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**76**	**76**	
Intercept	1	1	1	All levels present.
State	26	25	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	12	All levels present.
Total		**213**	**213**	

Exhibit H.1.3 Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 1: Northeast and South

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**86**	**86**	
Intercept	1	1	1	All levels present.
State	26	25	25	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**223**	**211**	

Exhibit H.1.4 Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 1: Northeast and South

This step used the same variables as the respondent pair-level poststratification step in Exhibit H.1.3.

Appendix H.2: Model Group 2: Midwest and West

(Alaska, Arizona, California, Colorado, Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Montana, Nebraska, Nevada, New Mexico, North Dakota, Ohio, Oregon, South Dakota, Utah, Washington, Wisconsin, Wyoming)

Table H.2a 2012 Pair Weight GEM Modeling Summary (Model Group 2: Midwest and West)

Modeling Step[1]	Extreme Weight Proportions			UWE[2]	# Covariates[3]	Bounds[4]	
	% Unweighted	% Weighted	% Winsorized			Nominal	Realized
sel.pr.ps	5.02	17.79	6.19	16.0574	212	(0.41, 2.00)	(0.42, 2.00)
	2.08	6.47	1.21	10.0819	200	(0.24, 3.46)	(0.25, 3.45)
						(0.90, 1.73)	(.0.90, 1.73)
res.pr.nr	2.31	9.13	1.90	10.7706	212	(1.01, 2.70)	(1.01, 2.70)
	2.87	11.04	2.21	11.1015	212	(1.00, 5.00)	(1.00, 5.00)
						N/A	N/A
res.pr.ps	2.83	12.68	3.78	11.1015	222	(0.56, 1.90)	(0.58, 1.90)
	2.25	12.52	2.14	9.2899	210	(0.43, 2.50)	(0.45, 2.49)
						N/A	N/A
res.pr.ev	2.25	12.52	2.14	9.2899	222	(0.80, 1.80)	(0.94, 1.77)
	1.04	7.73	0.48	8.7981	210	(0.70, 1.80)	(0.80, 1.30)
						N/A	N/A

GEM = generalized exponential model.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

[2] Unequal weighting effect (UWE) defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

[3] Number of proposed covariates on top line and number finalized after modeling.

[4] Nominal bounds are used in defining maximum/minimum values for the GEM adjustment factors. The realized bound is the actual adjustment produced by the modeling. The first set of bounds listed is for high extreme values, the second is for nonextreme values, and the third is for low extreme values.

Table H.2b 2012 Distribution of Weight Adjustment Factors and Weight Products (Model Group 2: Midwest and West)

	SDU Weight	Pair Selection		sel.pr.ps[1]		res.pr.nr[1]		res.pr.ps[1]		res.pr.ev[1]	
	1-10	pairwt11	1-11	pairwt12	1-12	pairwt13	1-13	pairwt14	1-14	pairwt15	1-15
Minimum	20	1.02	64	0.06	29	0.50	30	0.11	18	0.52	16
1%	92	1.10	161	0.31	132	0.99	153	0.50	124	0.83	118
5%	132	1.21	336	0.47	293	1.00	334	0.59	292	0.89	279
10%	175	1.32	550	0.58	470	1.01	518	0.66	474	0.92	462
25%	405	1.49	964	0.78	984	1.04	1,115	0.82	1,073	0.96	1,056
Median	577	5.26	2,692	1.01	2,564	1.15	2,913	0.97	2,888	0.99	2,861
75%	1,045	11.43	7,082	1.28	7,159	1.43	8,363	1.15	8,574	1.02	8,610
90%	1,691	21.68	15,760	1.59	17,689	1.93	22,988	1.45	23,090	1.06	23,367
95%	1,995	28.31	24,872	1.79	27,387	2.45	41,048	1.66	40,453	1.09	40,772
99%	2,664	54.34	57,015	2.42	67,482	4.09	118,207	2.03	122,494	1.14	121,848
Maximum	8,444	1,603.54	1,594,823	3.45	1,028,039	5.00	1,550,385	2.49	804,009	1.30	639,872
n	13,416	-	13,416	-	13,416	-	9,736	-	9,736	-	9,736
Mean	787	9.09	7,314	1.06	7,569	1.36	10,430	1.02	10,430	0.99	10,430
Max/Mean	11	-	218	-	136	-	149	-	77	-	61

SDU = screener dwelling unit.

[1] For a key to modeling abbreviations, see Chapter 6, Exhibit 6.1.

Model Group 2 Overview

Selected Pair-Level Poststratification

In the selected pair-level poststratification step, 200 of 212 proposed factors were retained in the final model. All main and two-factor effects were retained at proposed levels. None of the 12 three-factor effects were kept in the model due to convergence problems.

Respondent Pair-Level Nonresponse

In the respondent pair-level nonresponse step, all 212 proposed factors were retained in the final model.

Respondent Pair-Level Poststratification

In the respondent pair-level poststratification step, 210 of 222 proposed factors were retained in the final model, as in the selected pair-level poststratification step.

Respondent Pair-Level Extreme Value Adjustment

In the respondent pair-level extreme value adjustment step, 210 of 222 proposed factors were retained in the final model, as in the respondent pair-level poststratification step.

Exhibit H.2.1 Covariates for 2012 NSDUH Pair Weights (sel.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	25	24	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**212**	**200**	

Exhibit H.2.2 Covariates for 2012 NSDUH Pair Weights (res.pr.nr) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**75**	**75**	
Intercept	1	1	1	All levels present.
State	25	24	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**12**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	12	All levels present.
Total		**212**	**212**	

Exhibit H.2.3 Covariates for 2012 NSDUH Pair Weights (res.pr.ps) Model Group 2: Midwest and West

Variables	Level	Proposed	Final	Comments
One-Factor Effects		**85**	**85**	
Intercept	1	1	1	All levels present.
State	26	24	24	All levels present.
Quarter	4	3	3	All levels present.
Population Density	4	3	3	All levels present.
Group Quarter	3	2	2	All levels present.
Household Size	3	2	2	All levels present.
Pair Age	15	14	14	All levels present.
Pair Gender	3	2	2	All levels present.
Pair Race/Ethnicity	10	9	9	All levels present.
Race/Ethnicity of Householder	6	5	5	All levels present.
Rent/Housing	5	4	4	All levels present.
Segment % Black or African American	3	2	2	All levels present.
Segment % Hispanic or Latino	3	2	2	All levels present.
% Owner-Occupied	3	2	2	All levels present.
Pair Relationship	10	10	10	All levels present.
Two-Factor Effects		**125**	**125**	
Pair Race/Ethnicity (5 Levels) × Pair Age (6 Levels)	5 × 6	20	20	All levels present.
Pair Race/Ethnicity (5 Levels) × Pair Gender	5 × 3	8	8	All levels present.
Pair Gender × Pair Age (6 Levels)	3 × 6	10	10	All levels present.
State/Region × Pair Race/Ethnicity (5 Levels)	6 × 5	20	20	All levels present.
State/Region × Pair Age (6 Levels)	6 × 6	25	25	All levels present.
State/Region × Pair Gender	6 × 3	10	10	All levels present.
Rent/Housing × % Black or African American	5 × 3	8	8	All levels present.
Rent/Housing × % Hispanic or Latino	5 × 3	8	8	All levels present.
Rent/Housing × % Owner-Occupied	5 × 3	8	8	All levels present.
% Owner-Occupied × % Black or African American	3 × 3	4	4	All levels present.
% Owner-Occupied × % Hispanic or Latino	3 × 3	4	4	All levels present.
Three-Factor Effects		**12**	**0**	
Pair Race/Ethnicity (4 Levels) × Pair Gender × Pair Age (3 Levels)	4 × 3 × 3	12	0	Drop all; conv.
Total		**222**	**210**	

Exhibit H.2.4 Covariates for 2012 NSDUH Pair Weights (res.pr.ev) Model Group 2: Midwest and West

This step used the same variables as the respondent pair-level poststratification step in Exhibit H.2.3.

Appendix I: Evaluation of Calibration Weights: Pair-Level Response Rates

Table I.1 2012 NSDUH Person Pair-Level Response Rates

Domain	Selected Pairs	Respondent Pairs	% Interview Response Rate[1]
Total	27,035	19,459	63.76
Pair Age Group			
12-17, 12-17	4,507	3,666	81.51
12-17, 18-25	3,627	2,778	77.17
12-17, 26-34	825	653	77.42
12-17, 35-49	3,813	2,816	73.86
12-17, 50+	851	627	74.42
18-25, 18-25	5,476	3,976	73.18
18-25, 26-34	1,079	742	67.14
18-25, 35-49	1,582	1,058	65.84
18-25, 50+	1,074	650	58.92
26-34, 26-34	880	597	71.06
26-34, 35-49	469	305	52.70
26-34, 50+	315	178	53.89
35-49, 35-49	833	489	61.19
35-49, 50+	466	266	50.91
50+, 50+	1,238	658	52.22
Pair Race/Ethnicity			
Hispanic or Latino	4,176	3,084	69.40
Black or African American	2,760	2,210	71.36
White	16,039	11,284	61.92
Other	1,791	1,175	50.45
White & Black or African American	217	166	68.55
White & Hispanic or Latino	921	692	71.01
White & Other	737	552	63.97
Black or African American & Hispanic or Latino	123	91	69.54
Black or African American & Other	126	99	37.40
Hispanic or Latino & Other	145	106	58.42
Pair Gender			
Male, Male	5,928	4,173	60.83
Female, Female	5,813	4,425	67.33
Male, Female	15,294	10,861	63.58
Household Size			
Two	6,855	4,660	59.76
Three	7,398	5,248	60.52
Four or More	12,782	9,551	67.43

Table I.1 2012 NSDUH Person Pair-Level Response Rates (continued)

Domain	Selected Pairs	Respondent Pairs	% Interview Response Rate[1]
Census Region			
Northeast	5,685	3,856	58.61
South	7,934	5,867	65.95
Midwest	7,515	5,455	66.14
West	5,901	4,281	62.52
Quarter			
Quarter 1	6,334	4,613	64.66
Quarter 2	7,183	5,115	63.01
Quarter 3	7,012	5,093	64.85
Quarter 4	6,506	4,638	62.52
% Hispanic or Latino in Segment			
50-100%	2,105	1,537	65.52
10-<50%	5,112	3,726	64.88
<10%	19,818	14,196	63.07
% Black or African American in Segment			
50-100%	1,905	1,482	64.05
10-<50%	4,030	2,974	66.88
<10%	21,100	15,003	63.01
% Owner-Occupied DUs in Segment			
50-100%	20,778	14,866	63.43
10-<50%	4,889	3,600	64.61
<10%	1,368	993	71.03
Combined Median Rent/Housing Value			
1st Quintile	4,264	3,227	68.88
2nd Quintile	5,758	4,250	67.80
3rd Quintile	6,136	4,364	65.47
4th Quintile	6,013	4,276	58.78
5th Quintile	4,864	3,342	60.58
Population Density			
Large MSA	11,649	8,192	61.00
Medium to Small MSA	13,392	9,833	67.16
Non-MSA, Urban	511	380	71.70
Non-MSA, Rural	1,483	1,054	65.73
Group Quarters			
Group	386	312	79.80
Non-Group	26,649	19,147	63.70

DU = dwelling unit, MSA = metropolitan statistical area.

[1] The weight used for calculating the response rate includes screener dwelling unit (SDU)- and pair-level design weights, SDU nonresponse and poststratification adjustments, and selected pair poststratification adjustment. This weight is the product of WT1*…*WT9*PRWT10*PRWT11.

Appendix J: Evaluation of Calibration Weights: Pair-Level Proportions of Extreme Values and Outwinsors

Table J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.pr.ps[1] (SDUWT*PRWT10)			After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	27,035	2.14	5.22	1.44	5.34	26.13	15.62	2.15	15.29	7.51
Pair Age Group										
12-17, 12-17	4,507	1.84	4.16	1.29	3.64	17.26	6.55	0.55	3.06	0.64
12-17, 18-25	3,627	2.21	5.76	1.79	8.08	28.38	11.52	1.96	7.45	1.53
12-17, 26-34	825	1.82	4.03	1.52	1.82	5.22	1.36	0.85	2.56	0.36
12-17, 35-49	3,813	1.60	3.70	0.89	2.12	9.32	2.46	0.87	3.65	0.67
12-17, 50+	851	2.12	6.13	1.48	2.23	10.42	3.28	0.59	0.89	0.17
18-25, 18-25	5,476	2.12	4.99	1.22	8.16	27.90	10.87	3.67	12.83	2.39
18-25, 26-34	1,079	3.61	7.88	1.97	3.15	13.23	4.97	2.32	8.97	1.53
18-25, 35-49	1,582	3.41	9.38	3.19	6.76	26.19	11.00	2.40	7.55	1.43
18-25, 50+	1,074	2.33	5.87	1.61	3.82	16.98	6.23	1.30	3.90	0.59
26-34, 26-34	880	2.73	5.15	1.11	3.64	21.67	12.88	2.84	21.62	10.40
26-34, 35-49	469	2.77	5.47	1.26	7.46	43.03	29.62	5.97	36.72	19.45
26-34, 50+	315	2.22	5.94	2.07	4.44	22.48	12.39	1.59	16.78	12.07
35-49, 35-49	833	2.04	6.54	2.01	6.24	44.99	30.94	4.80	45.93	30.71
35-49, 50+	466	1.50	4.26	0.59	1.93	20.54	15.31	2.15	21.08	13.47
50+, 50+	1,238	1.62	2.89	0.40	8.24	43.73	33.58	4.36	21.24	8.98
Pair Race/Ethnicity										
Hispanic or Latino	4,176	3.86	10.45	3.98	5.17	35.09	24.22	2.95	22.46	11.57
Black or African American	2,760	3.44	5.97	1.01	6.63	33.29	22.33	3.22	20.24	10.19
White	16,039	0.68	1.49	0.19	4.65	21.19	12.31	1.35	11.75	6.17
Other	1,791	6.03	12.30	2.98	7.76	28.26	10.20	3.35	15.98	5.08
White & Black or African American	217	6.45	9.80	1.92	9.68	16.84	6.21	13.82	19.84	3.28
White & Hispanic or Latino	921	2.71	5.44	1.05	6.84	21.82	7.40	3.04	11.85	3.66
White & Other	737	3.53	6.72	1.57	4.75	18.77	8.09	2.04	14.84	7.83
Black or African American & Hispanic or Latino	123	17.89	43.36	18.62	17.89	66.55	44.19	6.50	39.53	17.08
Black or African American & Other	126	5.56	12.08	2.35	5.56	40.75	27.57	1.59	34.09	21.52
Hispanic or Latino & Other	145	8.28	16.29	5.15	8.97	21.25	9.63	6.90	7.90	1.71
Pair Gender										
Male, Male	5,928	2.14	4.77	1.29	7.25	31.61	20.02	2.65	11.86	5.08
Female, Female	5,813	2.34	5.51	1.66	5.33	18.19	7.47	2.31	11.75	3.90
Male, Female	15,294	2.07	5.28	1.42	4.61	26.61	16.48	1.90	17.24	9.21
Household Size										
Two	6,855	1.74	4.34	1.11	0.83	2.47	0.67	0.44	2.29	0.31
Three	7,398	2.10	5.32	1.69	2.61	27.75	19.75	1.82	16.75	7.66
Four or More	12,782	2.39	5.59	1.47	9.35	37.59	21.18	3.25	20.99	11.01

Table J.1 2012 NSDUH Selected Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)			Before sel.pr.ps[1] (SDUWT*PRWT10)			After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	5,685	2.62	6.98	2.28	5.82	21.94	9.44	2.36	15.05	6.33
South	7,934	1.78	4.49	1.05	5.29	30.24	20.63	1.79	15.23	7.69
Midwest	7,515	1.77	3.61	0.66	5.59	26.11	14.87	2.13	14.43	7.58
West	5,901	2.64	6.43	2.11	4.64	22.62	12.66	2.46	16.23	8.07
Quarter										
Quarter 1	6,334	2.07	4.60	1.28	5.32	29.27	20.03	2.68	13.34	5.68
Quarter 2	7,183	2.03	5.25	1.58	4.82	18.50	7.62	2.05	11.60	3.85
Quarter 3	7,012	2.05	4.87	1.15	5.03	25.81	14.78	1.64	19.10	10.98
Quarter 4	6,506	2.43	6.14	1.77	6.29	30.19	19.30	2.29	17.08	9.51
% Hispanic or Latino in Segment										
50-100%	2,105	2.28	5.26	1.89	4.42	36.21	27.31	2.23	17.79	8.80
10-<50%	5,112	3.66	9.15	3.09	5.97	33.28	22.09	3.01	18.47	9.83
<10%	19,818	1.74	3.74	0.76	5.28	21.35	10.80	1.92	13.74	6.47
% Black or African American in Segment										
50-100%	1,905	3.67	8.74	2.71	6.72	27.02	13.67	3.78	19.57	8.12
10-<50%	4,030	2.68	6.54	2.12	6.23	36.11	25.20	2.70	16.10	7.76
<10%	21,100	1.90	4.52	1.14	5.05	23.47	13.34	1.90	14.68	7.40
% Owner-Occupied DUs[1] in Segment										
50-100%	20,778	1.90	4.79	1.30	5.09	27.19	16.70	2.19	15.79	7.99
10-<50%	4,889	2.88	6.37	1.88	6.14	22.44	12.07	2.39	14.14	5.88
<10%	1,368	3.14	7.07	1.97	6.36	18.81	7.21	0.66	2.43	0.34
Combined Median Rent/Housing Value										
1st Quintile	4,264	1.74	4.47	1.35	5.42	39.11	29.10	2.11	20.67	12.05
2nd Quintile	5,758	2.15	5.91	1.99	5.45	22.10	9.42	2.57	16.93	7.54
3rd Quintile	6,136	2.22	4.62	1.18	5.30	23.81	13.29	1.65	14.94	8.36
4th Quintile	6,013	2.33	5.21	1.39	5.32	26.61	17.44	2.06	14.47	7.09
5th Quintile	4,864	2.16	5.64	1.28	5.24	20.76	10.26	2.43	11.49	4.17
Population Density										
Large MSA[1]	11,649	2.83	6.69	1.90	5.55	27.46	16.79	2.59	16.62	8.29
Medium to Small MSA[1]	13,392	1.68	3.67	0.97	5.39	19.31	7.73	1.92	12.88	5.52
Non-MSA,[1] Urban	511	0.78	0.68	0.13	3.33	72.80	70.94	1.37	23.83	19.61
Non-MSA,[1] Rural	1,483	1.35	2.73	0.51	3.98	26.19	19.15	1.01	17.82	12.43
Group Quarters										
Group	386	2.59	3.81	0.78	8.03	23.60	6.46	6.99	19.26	3.27
Non-Group	26,649	2.14	5.23	1.45	5.31	26.14	15.65	2.08	15.28	7.53

[1] This step used demographic variables from screener data for all selected person pairs; DU = dwelling unit, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, SDU = screener dwelling unit, Sel = selected.

[2] Weighted extreme value proportion: $100*\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values, and w_k denotes the weight for both extreme values and nonextreme values.

[3] Outwinsor weight proportion: $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table J.2 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)			After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	19,459	2.16	15.29	7.43	2.61	15.86	6.50
Pair Age Group							
12-17, 12-17	3,666	0.46	2.36	0.35	0.38	3.16	0.97
12-17, 18-25	2,778	2.12	8.17	1.63	1.73	7.94	2.34
12-17, 26-34	653	1.07	4.90	0.74	1.38	7.70	1.95
12-17, 35-49	2,816	0.96	4.78	0.78	1.07	2.89	0.55
12-17, 50+	627	0.32	0.72	0.22	0.64	3.03	0.66
18-25, 18-25	3,976	3.65	13.28	2.63	4.45	17.62	3.91
18-25, 26-34	742	2.96	11.93	1.64	4.04	17.04	3.72
18-25, 35-49	1,058	2.46	7.42	1.52	4.73	15.74	3.50
18-25, 50+	650	1.23	2.95	0.54	2.00	5.21	0.72
26-34, 26-34	597	3.85	25.89	13.06	2.68	27.33	11.77
26-34, 35-49	305	6.89	29.07	11.40	10.49	34.39	14.18
26-34, 50+	178	2.25	20.31	12.72	1.69	10.63	7.06
35-49, 35-49	489	5.32	49.78	35.62	7.77	48.74	30.19
35-49, 50+	266	3.38	26.20	12.42	4.14	16.72	6.22
50+, 50+	658	3.80	21.67	11.23	4.86	17.34	6.29
Pair Race/Ethnicity							
Hispanic or Latino	3,084	2.95	26.88	14.07	3.44	23.61	11.28
Black or African American	2,210	3.21	18.30	8.75	2.08	15.61	7.34
White	11,284	1.27	10.54	5.41	2.03	12.64	5.04
Other	1,175	3.32	18.31	5.76	6.47	26.06	7.61
White & Black or African American	166	14.46	20.29	4.46	10.84	15.25	3.46
White & Hispanic or Latino	692	3.76	13.37	4.64	0.87	7.53	4.25
White & Other	552	1.99	6.27	1.52	1.09	13.41	2.08
Black or African American & Hispanic or Latino	91	5.49	29.46	19.46	7.69	27.76	12.80
Black or African American & Other	99	2.02	8.26	0.30	10.10	20.77	7.12
Hispanic or Latino & Other	106	8.49	12.57	2.94	2.83	10.03	1.10
Pair Gender							
Male, Male	4,173	2.64	10.72	3.25	3.50	15.22	4.05
Female, Female	4,425	2.31	8.63	2.15	2.10	9.70	2.24
Male, Female	10,861	1.92	18.50	10.11	2.47	17.77	8.38
Household Size							
Two	4,660	0.45	2.58	0.43	0.67	4.13	1.44
Three	5,248	1.91	17.38	8.33	2.76	19.63	7.37
Four or More	9,551	3.14	19.92	10.09	3.47	19.72	8.56

Table J.2 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)			After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region							
Northeast	3,856	2.26	14.23	6.16	2.52	17.44	6.83
South	5,867	1.62	13.45	7.09	2.01	13.98	5.98
Midwest	5,455	2.11	15.39	8.50	2.88	19.13	8.55
West	4,281	2.90	18.66	7.89	3.15	14.76	5.35
Quarter							
Quarter 1	4,613	2.64	13.92	5.48	3.14	12.82	4.46
Quarter 2	5,115	2.09	10.20	3.27	2.64	13.67	3.28
Quarter 3	5,093	1.45	18.44	11.32	1.65	18.64	9.28
Quarter 4	4,638	2.54	18.57	9.55	3.08	18.26	8.94
% Hispanic or Latino in Segment							
50-100%	1,537	2.93	21.09	10.38	2.80	17.33	7.53
10-<50%	3,726	3.46	21.35	10.55	3.65	19.43	8.57
<10%	14,196	1.74	12.08	5.77	2.31	14.34	5.58
% Black or African American in Segment							
50-100%	1,482	3.71	15.16	4.95	3.78	15.80	4.57
10-<50%	2,974	3.19	19.03	9.57	3.30	17.37	7.61
<10%	15,003	1.81	14.40	7.16	2.35	15.52	6.44
% Owner-Occupied DUs[1] in Segment							
50-100%	14,866	2.07	15.63	7.94	2.52	15.96	6.65
10-<50%	3,600	2.86	14.94	5.66	3.36	15.87	6.06
<10%	993	1.01	4.00	0.63	1.21	11.38	3.21
Combined Median Rent/Housing Value							
1st Quintile	3,227	2.17	22.21	14.19	2.14	23.88	13.52
2nd Quintile	4,250	2.49	18.39	8.55	2.49	18.83	7.49
3rd Quintile	4,364	1.81	16.57	8.97	2.47	17.25	7.64
4th Quintile	4,276	2.03	9.98	2.44	3.04	10.26	2.39
5th Quintile	3,342	2.36	11.42	5.10	2.81	12.86	4.56
Population Density							
Large MSA[1]	8,192	2.70	16.16	7.80	3.33	15.99	6.41
Medium to Small MSA[1]	9,833	1.88	13.18	5.37	2.16	14.08	5.06
Non-MSA,[1] Urban	380	1.58	33.21	27.29	1.84	33.42	22.73
Non-MSA,[1] Rural	1,054	0.85	18.99	16.13	1.42	25.72	16.42
Group Quarters							
Group	312	8.33	25.50	3.58	5.77	23.85	6.74
Non-Group	19,147	2.06	15.25	7.44	2.55	15.83	6.50

[1] This step used demographic variables from screener data for all responding person pairs; DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PR = pair, Res = respondent, SDU = screener dwelling unit.

[2] Weighted extreme value proportion: $100*\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values, and w_k denotes the weight for both extreme values and nonextreme values.

[3] Outwinsor weight proportion: $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

Table J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Total	19,459	2.46	10.90	2.71	1.86	8.68	1.28	0.93	5.01	0.33
Pair Age Group										
12-17, 12-17	3,668	0.41	3.31	1.07	0.41	1.82	0.25	0.27	1.52	0.16
12-17, 18-25	2,759	1.70	7.95	2.36	0.91	2.82	0.58	0.69	2.47	0.36
12-17, 26-34	658	1.37	7.58	1.92	1.22	3.40	0.26	0.61	1.83	0.12
12-17, 35-49	2,812	1.07	2.84	0.53	1.21	3.50	0.58	0.71	2.08	0.20
12-17, 50+	631	0.63	3.04	0.65	0.95	5.54	1.15	0.48	2.34	0.23
18-25, 18-25	3,901	4.43	17.14	3.78	3.05	10.07	1.19	0.95	3.40	0.24
18-25, 26-34	794	3.78	18.47	7.16	3.78	13.35	1.57	1.76	5.64	0.41
18-25, 35-49	1,053	5.03	16.30	3.45	5.13	14.84	2.08	2.28	6.87	0.62
18-25, 50+	660	1.97	5.13	0.75	2.58	10.37	1.51	1.52	3.63	0.23
26-34, 26-34	604	1.66	10.16	3.30	0.50	7.55	1.26	0.50	7.77	0.36
26-34, 35-49	320	8.13	15.90	5.30	1.25	3.16	0.33	0.94	2.71	0.12
26-34, 50+	177	2.26	13.69	2.88	0.56	5.68	0.28	0.00	0.00	0.00
35-49, 35-49	487	5.95	24.71	8.32	1.23	7.32	1.46	0.82	5.68	0.19
35-49, 50+	272	2.57	12.66	2.81	1.84	9.61	1.24	0.74	3.39	0.18
50+, 50+	663	4.37	9.62	1.32	5.28	14.60	2.27	4.07	11.55	0.69
Pair Race/Ethnicity										
Hispanic or Latino	3,134	3.06	16.96	3.48	1.79	9.82	1.35	1.40	5.41	0.52
Black or African American	2,139	2.10	7.08	1.49	1.92	6.00	0.63	0.70	1.36	0.14
White	11,006	1.92	8.98	2.34	1.33	6.65	1.04	0.31	3.11	0.15
Other	1,116	6.36	22.29	6.39	6.72	25.53	3.65	5.73	22.66	1.50
White & Black or African American	167	8.38	14.47	3.01	10.78	19.96	3.60	5.99	11.36	0.65
White & Hispanic or Latino	721	1.25	2.96	0.51	1.94	11.77	1.94	0.55	8.94	0.15
White & Other	715	0.98	2.51	0.70	0.28	0.80	0.02	0.00	0.00	0.00
Black or African American & Hispanic or Latino	109	9.17	15.51	7.09	5.50	6.74	1.02	7.34	7.51	0.66
Black or African American & Other	198	6.06	23.35	8.75	1.52	8.09	2.89	0.51	5.37	1.63
Hispanic or Latino & Other	154	2.60	1.72	0.79	0.65	1.00	0.16	0.00	0.00	0.00
Pair Gender										
Male, Male	4,162	3.41	13.03	3.17	2.33	7.66	1.10	0.96	3.83	0.25
Female, Female	4,430	2.19	9.64	2.19	1.81	7.70	1.10	0.93	4.16	0.44
Male, Female	10,867	2.21	10.65	2.73	1.70	9.23	1.38	0.91	5.57	0.31
Household Size										
Two	4,660	0.64	3.96	1.39	0.43	1.72	0.21	0.34	1.53	0.14
Three	5,248	2.69	15.06	4.04	2.02	13.08	2.13	1.28	9.73	0.61
Four or More	9,551	3.22	12.17	2.68	2.47	9.83	1.36	1.02	4.27	0.27

Table J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Census Region										
Northeast	3,856	2.54	12.15	2.44	1.92	7.94	0.95	1.32	5.11	0.39
South	5,867	1.79	7.96	1.49	1.18	4.14	0.35	0.48	1.46	0.09
Midwest	5,455	2.64	14.44	4.77	2.42	11.90	2.00	1.06	6.70	0.47
West	4,281	3.08	11.28	2.99	2.03	13.00	2.25	1.00	8.56	0.49
Quarter										
Quarter 1	4,613	2.99	8.43	2.00	1.99	6.66	1.02	0.89	4.69	0.30
Quarter 2	5,115	2.35	9.58	1.92	1.62	7.03	1.10	0.92	3.44	0.28
Quarter 3	5,093	1.49	9.89	3.06	1.57	12.63	1.87	0.98	7.23	0.46
Quarter 4	4,638	3.13	15.67	3.86	2.31	8.33	1.11	0.91	4.65	0.26
% Hispanic or Latino in Segment										
50-100%	1,537	2.47	8.16	1.91	1.63	7.42	1.05	1.11	4.54	0.30
10-<50%	3,726	3.38	14.96	3.61	2.74	13.61	1.92	1.83	8.96	0.67
<10%	14,196	2.22	9.86	2.52	1.66	7.09	1.08	0.67	3.65	0.21
% Black or African American in Segment										
50-100%	1,482	3.58	11.09	2.83	2.50	8.02	0.97	1.62	3.27	0.34
10-<50%	2,974	3.03	10.73	2.90	2.35	9.89	1.39	1.68	5.52	0.35
<10%	15,003	2.24	10.91	2.66	1.70	8.46	1.28	0.71	5.06	0.32
% Owner-Occupied DUs[1] in Segment										
50-100%	14,866	2.36	10.81	2.54	1.59	8.49	1.26	0.73	4.76	0.30
10-<50%	3,600	3.25	11.32	3.53	2.86	9.19	1.26	1.53	5.55	0.39
<10%	993	1.11	11.07	3.09	2.32	12.84	2.43	1.61	11.60	1.12
Combined Median Rent/Housing Value										
1st Quintile	3,227	2.05	16.76	5.92	1.77	12.66	2.44	0.90	6.50	0.54
2nd Quintile	4,250	2.24	7.56	1.39	1.53	6.50	0.80	0.73	4.36	0.27
3rd Quintile	4,364	2.29	11.09	2.09	1.60	6.66	0.83	0.69	2.28	0.15
4th Quintile	4,276	2.92	9.78	2.64	2.27	7.94	1.11	1.05	4.95	0.28
5th Quintile	3,342	2.78	11.32	2.57	2.18	10.99	1.62	1.35	7.41	0.46
Population Density										
Large MSA[1]	8,192	3.15	12.57	2.97	2.45	9.78	1.33	1.39	5.05	0.36
Medium to Small MSA[1]	9,833	2.02	7.23	1.47	1.38	6.92	1.07	0.59	4.69	0.28
Non-MSA,[1] Urban	380	1.84	4.24	1.68	1.05	0.63	0.21	0.26	0.09	0.00
Non-MSA,[1] Rural	1,054	1.42	25.72	11.58	1.99	13.15	2.95	0.66	9.05	0.39
Group Quarters										
Group	312	6.73	23.62	6.88	8.97	20.47	3.20	2.88	10.63	0.39
Non-Group	19,147	2.39	10.85	2.70	1.74	8.63	1.27	0.89	4.99	0.33

Table J.3 2012 NSDUH Respondent Pair-Level Proportions of Extreme Values and Outwinsors (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)			After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)			Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)		
		% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]	% Unweighted	% Weighted[2]	% Outwinsor[3]
Pair Relationship Domain[4]										
Parent-Child (12-14)	1,981	0.91	3.04	0.71	1.31	3.74	0.58	0.56	1.53	0.11
Parent-Child (12-17)	3,748	0.96	3.49	0.74	1.17	3.83	0.57	0.59	1.89	0.18
Parent-Child (12-20)	4,443	1.67	6.67	1.28	2.21	8.89	1.33	1.04	3.48	0.34
Sibling (12-14)-Sibling (15-17)	2,191	0.46	4.07	1.43	0.46	2.39	0.36	0.32	2.01	0.24
Sibling (12-17)-Sibling (18-25)	2,477	1.53	7.04	1.82	0.89	2.81	0.54	0.61	2.11	0.31
Spouse-Spouse/Partner-Partner	3,664	2.07	9.93	2.85	1.83	8.06	1.30	0.98	6.75	0.31
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	1,640	1.77	14.25	3.18	2.87	8.20	1.34	0.91	5.10	0.15

[1] This step used demographic variables from questionnaire data for all responding person pairs; DU = dwelling unit, EV = extreme value adjustment, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, Res = respondent, SDU = screener dwelling unit.

[2] Weighted extreme value proportion: $100*\sum_k w_{ek}/\sum_k w_k$, where w_{ek} denotes the weight for extreme values, and w_k denotes the weight for both extreme values and nonextreme values.

[3] Outwinsor weight proportion: $100*\sum_k(w_{ek} - b_k)/\sum_k w_k$, where b_k denotes the winsorized weight.

[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain.

Appendix K: Evaluation of Calibration Weights: Pair-Level Slippage Rates

Table K.1 2012 NSDUH Respondent Pair-Level Slippage Rates

Domain	*n*	Initial Total (*I*)[1]	Final Total (*F*)[2]	Control Total from SDU (*C*)	(*I* - *C*)/*C*%	(*F* - *C*)/*C*%
Total	19,459	221,952,167	221,952,167	221,952,167	-0.00	-0.00
Pair Age Group						
12-17, 12-17	3,668	7,255,585	7,213,926	7,213,926	0.58	-0.00
12-17, 18-25	2,759	8,114,464	8,172,406	8,172,406	-0.71	-0.00
12-17, 26-34	658	5,360,122	5,281,913	5,281,913	1.48	-0.00
12-17, 35-49	2,812	29,558,177	29,592,182	29,592,182	-0.11	-0.00
12-17, 50+	631	11,948,454	11,999,450	11,999,450	-0.42	-0.00
18-25, 18-25	3,901	12,207,700	12,545,793	12,545,793	-2.69	-0.00
18-25, 26-34	794	8,056,376	7,391,400	7,391,400	9.00	-0.00
18-25, 35-49	1,053	17,787,086	17,720,889	17,720,889	0.37	-0.00
18-25, 50+	660	18,580,624	18,321,739	18,321,739	1.41	-0.00
26-34, 26-34	604	9,996,036	10,662,196	10,662,196	-6.25	0.00
26-34, 35-49	320	8,782,458	8,567,386	8,567,386	2.51	0.00
26-34, 50+	177	10,685,258	11,157,395	11,157,395	-4.23	0.00
35-49, 35-49	487	18,497,856	18,578,012	18,578,012	-0.43	0.00
35-49, 50+	272	16,664,510	16,640,517	16,640,517	0.14	-0.00
50+, 50+	663	38,457,461	38,106,962	38,106,962	0.92	0.00
Pair Race/Ethnicity						
Hispanic or Latino	3,134	38,960,071	38,601,525	38,601,525	0.93	-0.00
Black or African American	2,139	24,049,201	24,291,619	24,291,619	-1.00	-0.00
White	11,006	121,400,147	123,391,386	123,391,386	-1.61	0.00
Other	1,116	15,936,917	16,500,923	16,500,923	-3.42	-0.00
White & Black or African American	167	1,811,631	2,028,598	2,028,598	-10.70	0.00
White & Hispanic or Latino	721	7,863,668	8,002,876	8,002,876	-1.74	0.00
White & Other	715	6,519,769	5,364,414	5,364,414	21.54	0.00
Black or African American & Hispanic or Latino	109	1,375,212	1,561,670	1,561,670	-11.94	-0.00
Black or African American & Other	198	2,236,103	1,038,323	1,038,323	115.36	-0.00
Hispanic or Latino & Other	154	1,799,447	1,170,834	1,170,834	53.69	0.00
Pair Gender						
Male, Male	4,162	39,706,191	39,891,754	39,891,754	-0.47	-0.00
Female, Female	4,430	39,862,298	39,874,794	39,874,794	-0.03	-0.00
Male, Female	10,867	142,383,679	142,185,618	142,185,618	0.14	-0.00
Pair Relationship Domain[3,4,5]						
Parent-Child (12-14)*	1,981	11,386,620	12,505,760	12,505,760	-8.95	-0.00
Parent-Child (12-17)*	3,748	23,617,276	25,045,406	25,045,406	-5.70	-0.00
Parent-Child (15-17)*	1,767	12,230,656	12,539,646	12,539,646	-2.46	-0.00
Parent-Child (12-20)*	4,443	31,454,617	34,378,755	34,378,755	-8.51	-0.00
Parent*-Child (12-14)	1,981	18,350,450	19,342,709	19,342,709	-5.13	-0.00
Parent*-Child (12-17)	3,748	30,953,375	32,235,193	32,235,193	-3.98	-0.00
Parent*-Child (15-17)	1,767	19,026,718	19,334,414	19,163,807	-0.72	0.89
Parent*-Child (12-20)	4,443	38,356,007	40,347,812	40,347,812	-4.94	-0.00
Sibling (12-14)-Sibling (15-17)*	2,191	3,929,975	4,012,028	4,012,028	-2.05	-0.00
Sibling (12-17)-Sibling (18-25)*	2,477	5,978,196	6,172,680	6,172,680	-3.15	-0.00
Spouse-Spouse/Partner-Partner	3,664	70,208,199	71,421,020	71,421,020	-1.70	-0.00
Spouse-Spouse/Partner-Partner with Children (Younger Than 18)	1,640	22,125,791	29,600,927	29,600,927	-25.25	-0.00

Table K.1 2012 NSDUH Respondent Pair-Level Slippage Rates (continued)

Domain	*n*	Initial Total (*I*)[1]	Final Total (*F*)[2]	Control Total from SDU (*C*)	(*I* - *C*)/*C*%	(*F* - *C*)/*C*%
Household Size						
Two	4,660	54,622,573	54,622,573	54,622,573	-0.00	-0.00
Three	5,248	57,411,679	57,411,679	57,411,679	-0.00	-0.00
Four or More	9,551	109,917,914	109,917,914	109,917,914	-0.00	-0.00
Census Region						
Northeast	3,856	41,025,207	41,025,207	41,025,207	-0.00	-0.00
South	5,867	79,383,669	79,383,669	79,383,669	0.00	-0.00
Midwest	5,455	44,933,034	44,933,034	44,933,034	-0.00	-0.00
West	4,281	56,610,256	56,610,256	56,610,256	-0.00	-0.00
Quarter						
Quarter 1	4,613	54,726,611	54,726,611	54,726,611	-0.00	-0.00
Quarter 2	5,115	55,791,727	55,791,727	55,791,727	-0.00	-0.00
Quarter 3	5,093	55,933,469	55,933,469	55,933,469	-0.00	-0.00
Quarter 4	4,638	55,500,360	55,500,360	55,500,360	-0.00	-0.00
% Hispanic or Latino in Segment						
50-100%	1,537	23,192,230	23,192,230	23,192,230	-0.00	-0.00
10-<50%	3,726	52,879,054	52,879,055	52,879,055	-0.00	-0.00
<10%	14,196	145,880,883	145,880,883	145,880,883	-0.00	-0.00
% Black or African American in Segment						
50-100%	1,482	16,732,625	16,732,625	16,732,625	-0.00	-0.00
10-<50%	2,974	38,150,858	38,150,858	38,150,858	-0.00	-0.00
<10%	15,003	167,068,684	167,068,684	167,068,684	-0.00	-0.00
% Owner-Occupied DUs in Segment						
50-100%	14,866	182,125,104	182,125,104	182,125,104	-0.00	-0.00
10-<50%	3,600	35,952,245	35,952,245	35,952,245	-0.00	-0.00
<10%	993	3,874,817	3,874,817	3,874,817	-0.00	-0.00
Combined Median Rent/Housing Value						
1st Quintile	3,227	30,573,140	30,573,140	30,573,140	-0.00	-0.00
2nd Quintile	4,250	44,525,897	44,525,897	44,525,897	-0.00	-0.00
3rd Quintile	4,364	46,327,833	46,327,833	46,327,833	-0.00	-0.00
4th Quintile	4,276	53,795,269	53,795,269	53,795,269	-0.00	-0.00
5th Quintile	3,342	46,730,029	46,730,029	46,730,029	-0.00	-0.00
Population Density						
Large MSA	8,192	122,562,456	122,562,456	122,562,456	-0.00	-0.00
Medium to Small MSA	9,833	87,447,088	87,447,088	87,447,088	-0.00	-0.00
Non-MSA, Urban	380	2,868,028	2,868,028	2,868,028	0.00	0.00
Non-MSA, Rural	1,054	9,074,595	9,074,595	9,074,595	-0.00	-0.00
Group Quarters						
Group	312	798,082	798,082	798,082	0.00	0.00
Non-Group	19,147	221,154,085	221,154,085	221,154,085	-0.00	-0.00

DU = dwelling unit, MSA = metropolitan statistical area, SDU = screener dwelling unit.

[1] WT1*...*WT10*PRWT11*...*PRWT13 (before person pair poststratification).

[2] WT1*...*WT10*PRWT11*...*PRWT14 (after person pair poststratification).

[3] The member of the pair that is the focus is designated with an asterisk (*).

[4] The parent-child (15-17) pair domains were not controlled for within the modeling and thus have higher slippage rates than the other domains listed. However, since these domains are a subset of other controlled domains, the rates are not large.

[5] Slippage rates were not calculated for the sibling-sibling domains with the younger child as the focus since no household counts for this domain were calculated and are required to construct the appropriate controls totals.

Appendix L: Evaluation of Calibration Weights: Pair-Level Weight Summary Statistics

Table L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	27,035	9	441	678	1,134	8,444	1.58	20	1,136	2,941	7,922	6,068,764	46.74	8	1,023	2,791	7,871	1,053,229	9.27
Pair Age Group																			
12-17, 12-17	4,507	16	399	609	1,042	7,853	1.62	20	582	1,017	1,841	56,843	3.05	8	516	998	1,969	25,287	2.40
12-17, 18-25	3,627	18	476	694	1,182	8,444	1.60	26	744	1,363	2,544	52,622	2.84	15	709	1,451	2,854	31,529	2.22
12-17, 26-34	825	32	397	675	1,131	5,004	1.53	248	2,487	4,469	7,333	70,973	2.29	143	1,949	3,759	7,455	69,227	2.63
12-17, 35-49	3,813	20	435	626	1,064	8,351	1.52	141	2,952	5,274	8,974	168,153	2.41	114	2,321	4,657	9,259	117,119	2.55
12-17, 50+	851	56	472	727	1,164	7,097	1.56	878	5,875	10,166	17,187	219,397	2.32	229	4,637	9,115	17,385	156,917	2.33
18-25, 18-25	5,476	9	432	704	1,207	7,653	1.62	32	798	1,392	2,461	41,249	2.70	15	589	1,253	2,936	39,303	2.52
18-25, 26-34	1,079	19	477	738	1,208	5,421	1.57	164	2,798	4,588	7,820	179,640	3.04	105	1,819	3,421	8,319	139,322	3.17
18-25, 35-49	1,582	34	491	711	1,196	7,847	1.61	489	3,467	6,420	12,481	245,571	3.24	180	2,804	5,903	13,746	105,508	2.60
18-25, 50+	1,074	28	508	704	1,240	6,541	1.54	377	6,270	9,858	17,777	412,510	3.25	282	5,644	10,249	20,203	187,095	2.45
26-34, 26-34	880	20	442	711	1,134	5,065	1.49	334	5,246	8,233	13,604	789,953	7.71	167	3,867	6,865	12,454	313,383	5.05
26-34, 35-49	469	43	478	731	1,206	3,895	1.46	571	6,016	9,618	16,559	1,594,823	18.62	412	5,391	9,152	16,689	440,141	5.90
26-34, 50+	315	43	466	757	1,126	7,250	1.57	1,140	11,808	19,298	29,840	554,381	3.33	488	10,916	21,855	40,294	701,763	3.70
35-49, 35-49	833	36	445	649	1,142	6,883	1.59	553	5,602	9,452	15,385	1,317,081	13.17	439	5,576	10,612	18,055	1,053,229	11.42
35-49, 50+	466	59	446	666	1,098	3,202	1.48	1,471	9,255	15,553	27,455	858,997	5.46	1,037	9,888	20,007	39,458	868,188	4.56
50+, 50+	1,238	32	456	664	1,083	3,612	1.43	1,016	12,714	19,083	33,172	6,068,764	42.37	684	12,424	20,906	35,262	856,498	3.21
Pair Race/Ethnicity																			
Hispanic or Latino	4,176	9	493	815	1,390	8,444	1.61	23	1,375	3,155	8,373	6,068,764	102.68	16	1,168	3,013	8,416	1,053,229	11.77
Black or African American	2,760	42	579	796	1,251	5,724	1.45	59	1,341	3,285	8,689	1,594,823	23.31	17	1,240	3,266	8,011	586,052	9.23
White	16,039	38	428	636	1,031	4,310	1.50	54	1,066	2,834	7,640	5,298,149	39.50	28	988	2,654	7,599	1,028,039	9.12
Other	1,791	19	256	585	1,242	5,171	1.82	26	976	2,726	7,732	272,377	5.93	14	774	2,543	7,854	293,645	6.53
White & Black or African American	217	53	523	816	1,211	3,937	1.47	193	1,697	3,581	8,188	57,878	2.48	68	2,166	4,858	11,831	97,680	2.63
White & Hispanic or Latino	921	18	430	701	1,231	6,940	1.63	20	1,148	3,423	9,307	324,611	5.04	8	1,009	3,223	8,847	277,585	4.94
White & Other	737	20	255	535	981	5,026	1.72	91	1,014	2,737	8,205	306,037	6.09	29	777	2,428	7,257	453,420	8.99
Black or African American & Hispanic or Latino	123	38	696	1,166	1,904	7,579	1.81	39	1,931	5,048	10,354	789,953	18.15	27	1,805	4,597	9,866	285,003	7.52
Black or African American & Other	126	36	380	563	994	3,202	1.67	65	1,112	2,716	6,804	371,337	17.37	15	726	2,375	5,646	334,168	16.81
Hispanic or Latino & Other	145	20	263	559	1,137	5,421	1.91	85	942	2,697	7,865	72,609	3.69	120	1,026	3,407	8,538	88,347	3.58

Table L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

Domain	n	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Pair Gender																			
Male, Male	5,928	18	446	705	1,137	7,154	1.56	20	1,020	2,442	6,430	5,298,149	99.54	8	995	2,517	6,644	856,498	8.36
Female, Female	5,813	9	426	651	1,126	8,359	1.61	23	1,032	2,588	6,877	348,424	5.05	15	924	2,455	6,648	527,688	6.63
Male, Female	15,294	18	447	676	1,136	8,444	1.58	31	1,242	3,295	9,094	6,068,764	40.82	14	1,078	3,126	9,128	1,053,229	9.71
Household Size																			
Two	6,855	20	428	663	1,087	6,940	1.54	38	1,320	4,305	11,582	105,073	2.48	15	862	2,867	10,535	192,611	3.36
Three	7,398	18	447	677	1,117	8,444	1.58	20	1,270	3,134	6,962	5,298,149	79.82	8	1,109	3,072	7,076	1,028,039	10.54
Four or More	12,782	9	442	689	1,176	8,359	1.60	23	1,014	2,371	7,069	6,068,764	52.53	14	1,069	2,633	7,458	1,053,229	11.36
Census Region																			
Northeast	5,685	18	335	605	843	7,128	1.61	20	998	2,548	7,060	451,843	6.11	8	784	2,357	6,775	586,033	8.38
South	7,934	9	659	958	1,348	7,138	1.40	23	1,545	4,149	10,607	6,068,764	79.32	15	1,332	3,723	10,044	1,053,229	8.39
Midwest	7,515	33	451	548	728	4,474	1.36	64	896	2,504	6,270	1,594,823	23.29	29	960	2,354	5,998	1,028,039	11.16
West	5,901	20	281	749	1,548	8,444	1.74	68	1,124	2,913	9,026	1,246,856	11.04	35	1,015	2,955	9,102	868,188	8.75
Quarter																			
Quarter1	6,334	21	496	733	1,173	7,847	1.54	26	1,205	2,948	7,968	6,068,764	90.25	14	1,182	3,093	8,706	674,933	6.79
Quarter2	7,183	18	409	619	1,092	7,853	1.58	31	1,069	2,743	7,529	789,953	6.39	16	966	2,669	7,602	456,537	5.56
Quarter3	7,012	9	434	631	1,141	7,138	1.59	20	1,077	2,911	7,852	1,246,856	12.56	8	895	2,545	7,289	868,188	12.34
Quarter4	6,506	16	476	711	1,145	8,444	1.60	23	1,179	3,230	8,450	5,298,149	63.92	15	1,071	2,967	8,090	1,053,229	12.19
% Hispanic or Latino in Segment																			
50-100%	2,105	55	583	1,046	1,488	6,365	1.36	85	1,810	4,352	10,864	6,068,764	114.14	42	1,465	3,748	10,433	674,933	7.42
10-<50%	5,112	9	570	895	1,480	8,444	1.52	23	1,571	3,777	10,099	5,298,149	57.95	16	1,455	3,901	10,200	1,053,229	9.93
<10%	19,818	16	389	613	987	7,853	1.58	20	1,008	2,647	7,213	1,594,823	11.23	8	914	2,467	7,090	1,028,039	8.91
% Black or African American in Segment																			
50-100%	1,905	9	556	763	1,172	7,128	1.59	23	1,312	3,254	8,310	371,337	6.18	15	1,180	3,228	8,308	527,688	7.76
10-<50%	4,030	22	541	832	1,337	7,847	1.50	31	1,436	3,529	9,224	6,068,764	95.73	27	1,237	3,357	9,426	800,116	8.03
<10%	21,100	18	408	639	1,089	8,444	1.58	20	1,071	2,821	7,657	5,298,149	33.09	8	979	2,674	7,618	1,053,229	9.72
% Owner-Occupied DUs[1] in Segment																			
50-100%	20,778	18	441	663	1,116	8,444	1.57	20	1,155	3,052	8,163	6,068,764	52.77	8	1,117	3,059	8,514	1,053,229	9.35
10-<50%	4,889	16	429	715	1,145	7,154	1.57	23	1,076	2,680	7,260	1,594,823	15.44	15	914	2,535	7,070	586,033	7.27
<10%	1,368	9	475	826	1,415	8,351	1.63	29	1,079	2,343	6,980	179,640	3.88	15	409	1,123	2,923	66,093	4.36
Combined Median Rent/Housing Value																			
1st Quintile	4,264	19	392	609	966	8,351	1.62	26	991	2,568	6,844	6,068,764	131.76	15	864	2,304	6,422	1,028,039	15.85
2nd Quintile	5,758	18	411	641	1,108	8,444	1.68	20	1,058	2,749	7,573	573,547	6.64	8	959	2,665	7,281	586,052	8.64
3rd Quintile	6,136	9	389	617	1,036	6,940	1.60	29	1,050	2,804	7,334	1,246,856	13.44	15	868	2,487	6,999	1,053,229	10.97
4th Quintile	6,013	16	478	750	1,210	7,853	1.54	23	1,266	3,248	8,839	5,298,149	65.38	14	1,161	3,121	8,853	868,188	8.61
5th Quintile	4,864	25	530	809	1,264	8,359	1.44	31	1,361	3,320	9,477	1,594,823	11.59	15	1,314	3,596	10,130	701,763	5.48

Table L.1 2012 NSDUH Selected Pair-Level Weight Summary Statistics (continued)

Domain	*n*	SDU-Level Weights[1] (SDUWT: WT1*...*WT9)						Before sel.pr.ps[1] (SDUWT*PRWT10)						After sel.pr.ps[1] (SDUWT*PRWT10*PRWT11)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Population Density																			
Large MSA[1]	11,649	9	586	869	1,395	8,359	1.42	23	1,593	3,974	10,354	5,298,149	33.22	15	1,538	3,988	10,441	1,053,229	8.13
Medium to Small MSA[1]	13,392	18	319	561	914	8,444	1.66	20	920	2,352	6,521	573,547	6.09	8	783	2,176	6,263	736,748	8.04
Non-MSA,[1] Urban	511	20	223	492	798	3,628	1.71	79	787	2,016	5,989	6,068,764	256.70	31	607	1,591	4,576	674,933	30.68
Non-MSA,[1] Rural	1,483	20	190	472	829	4,220	1.77	64	720	1,878	5,562	1,317,081	35.03	35	597	1,693	5,171	1,028,039	23.96
Group Quarters																			
Group	386	25	235	422	934	3,853	1.99	38	455	983	1,865	21,866	3.52	15	349	977	2,295	23,423	3.36
Non-Group	26,649	9	446	680	1,136	8,444	1.58	20	1,155	2,984	8,033	6,068,764	46.38	8	1,042	2,843	7,985	1,053,229	9.20

[1] This step used demographic variables from screener data for all selected person pairs; DU = dwelling unit, MSA = metropolitan statistical area, PR = pair, PS = poststratification, SDU = screener dwelling unit, Sel = selected.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Table L.2 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr)

Domain	*n*	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)						After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Total	19,459	8	965	2,556	7,018	1,053,229	9.79	10	1,129	3,171	9,270	1,550,385	10.18
Pair Age Group													
12-17, 12-17	3,666	8	515	1,007	1,973	17,520	2.32	10	590	1,169	2,375	34,664	2.58
12-17, 18-25	2,778	15	716	1,477	2,953	21,346	2.17	15	854	1,813	3,689	37,794	2.39
12-17, 26-34	653	143	1,901	3,700	7,161	69,227	2.69	143	2,053	4,448	8,869	151,613	3.23
12-17, 35-49	2,816	174	2,326	4,648	9,277	117,119	2.59	176	2,867	5,831	12,576	148,919	2.68
12-17, 50+	627	229	4,611	8,883	17,324	156,917	2.42	258	5,329	10,817	22,909	230,684	2.70
18-25, 18-25	3,976	15	587	1,252	2,936	39,303	2.56	17	645	1,528	3,893	44,826	2.84
18-25, 26-34	742	105	1,745	3,332	8,192	91,433	3.02	113	1,973	3,892	10,805	186,312	3.68
18-25, 35-49	1,058	180	2,675	5,658	13,746	105,508	2.64	183	3,407	8,114	21,008	184,727	2.83
18-25, 50+	650	282	5,534	9,839	19,855	148,367	2.33	289	7,718	15,800	34,492	296,605	2.53
26-34, 26-34	597	396	4,040	6,960	12,877	313,383	5.22	402	4,375	7,653	15,191	481,267	6.43
26-34, 35-49	305	412	4,837	8,543	15,131	382,959	4.50	412	7,194	13,271	29,726	776,131	5.09
26-34, 50+	178	989	10,601	21,015	38,170	701,763	4.02	991	12,816	31,414	78,144	919,088	3.14
35-49, 35-49	489	439	5,391	10,342	18,217	1,053,229	13.26	463	7,345	14,279	29,236	1,550,385	10.91
35-49, 50+	266	1,037	9,310	17,241	34,884	800,116	4.44	1,038	12,826	28,815	75,129	1,024,681	3.17
50+, 50+	658	753	13,107	20,833	34,426	674,933	3.29	1,612	20,499	36,877	69,712	876,058	2.68
Pair Race/Ethnicity													
Hispanic or Latino	3,084	22	1,110	2,795	7,446	1,053,229	14.92	25	1,237	3,235	9,954	1,323,156	13.93
Black or African American	2,210	21	1,175	3,050	7,271	586,052	8.76	21	1,270	3,630	8,954	876,058	9.62
White	11,284	28	927	2,403	6,635	1,028,039	8.77	38	1,107	3,054	8,977	1,550,385	9.94
Other	1,175	14	641	2,074	6,224	249,217	6.58	14	883	2,804	9,328	337,536	6.94
White & Black or African American	166	68	2,086	4,813	11,751	58,915	2.20	86	2,298	6,930	16,409	77,753	2.55
White & Hispanic or Latino	692	8	1,010	3,072	8,420	277,585	5.25	10	1,047	3,268	10,111	481,267	7.61
White & Other	552	29	760	2,335	6,975	118,112	4.53	29	816	2,564	8,087	355,652	8.32
Black or African American & Hispanic or Latino	91	27	1,805	4,597	9,157	285,003	8.23	27	1,965	5,172	12,415	285,060	6.03
Black or African American & Other	99	15	722	2,323	4,297	28,574	2.69	17	1,070	4,113	11,659	106,101	3.28
Hispanic or Latino & Other	106	120	1,132	3,393	6,819	37,041	2.69	126	1,363	4,407	10,976	104,183	3.84
Pair Gender													
Male, Male	4,173	8	941	2,356	5,891	342,223	5.71	10	1,130	2,898	8,040	488,930	7.38
Female, Female	4,425	15	882	2,362	6,170	249,217	4.74	17	1,024	2,850	7,945	285,780	6.13
Male, Female	10,861	14	1,012	2,769	7,844	1,053,229	11.30	14	1,170	3,457	10,446	1,550,385	11.15
Household Size													
Two	4,660	15	789	2,337	8,759	192,611	3.54	17	856	2,654	11,109	292,988	5.19
Three	5,248	8	992	2,765	6,318	1,028,039	11.78	10	1,189	3,534	8,651	1,550,385	12.87
Four or More	9,551	14	1,046	2,514	6,792	1,053,229	11.40	14	1,264	3,187	8,995	1,323,156	11.36

Table L.2 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.nr) (continued)

Domain	n	Before res.pr.nr[1] (SDUWT*PRWT10*PRWT11)						After res.pr.nr[1] (SDUWT*PRWT10*...*PRWT12)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Census Region													
Northeast	3,856	8	691	2,056	5,591	586,033	8.98	10	813	2,697	8,171	1,124,135	10.75
South	5,867	15	1,238	3,355	8,957	1,053,229	8.65	17	1,403	4,051	11,527	1,323,156	8.68
Midwest	5,455	29	922	2,184	5,400	1,028,039	13.98	30	1,083	2,645	6,808	1,550,385	14.22
West	4,281	35	986	2,738	8,023	800,116	8.29	36	1,161	3,392	10,802	1,024,681	8.68
Quarter													
Quarter1	4,613	14	1,138	2,868	7,787	674,933	6.90	14	1,351	3,582	10,437	854,597	7.22
Quarter2	5,115	22	912	2,501	6,737	285,003	4.94	25	1,071	3,131	8,922	417,070	6.38
Quarter3	5,093	8	830	2,311	6,395	800,116	12.88	10	963	2,769	8,219	1,024,681	12.79
Quarter4	4,638	15	1,000	2,635	7,081	1,053,229	14.19	15	1,189	3,328	9,797	1,550,385	14.07
% Hispanic or Latino in Segment													
50-100%	1,537	42	1,399	3,267	8,917	674,933	8.97	42	1,594	4,086	11,989	854,597	9.02
10-<50%	3,726	34	1,311	3,552	8,987	1,053,229	11.03	35	1,482	4,402	11,789	1,323,156	10.37
<10%	14,196	8	860	2,272	6,236	1,028,039	8.79	10	1,024	2,844	8,288	1,550,385	9.90
% Black or African American in Segment													
50-100%	1,482	15	1,076	2,881	7,396	342,223	5.53	17	1,232	3,602	10,074	488,930	6.27
10-<50%	2,974	29	1,178	3,074	8,215	800,116	9.87	29	1,287	3,799	10,451	1,024,681	9.57
<10%	15,003	8	923	2,428	6,752	1,053,229	10.14	10	1,089	3,042	8,915	1,550,385	10.70
% Owner-Occupied DUs[1] in Segment													
50-100%	14,866	8	1,055	2,758	7,528	1,053,229	9.93	10	1,248	3,505	10,121	1,550,385	9.97
10-<50%	3,600	15	855	2,336	6,170	586,033	7.50	17	991	2,731	8,123	1,124,135	9.53
<10%	993	15	419	1,065	2,810	66,093	4.60	17	458	1,220	3,258	133,448	6.99
Combined Median Rent/Housing Value													
1st Quintile	3,227	15	853	2,199	5,701	1,028,039	20.71	15	955	2,526	6,979	1,550,385	21.52
2nd Quintile	4,250	8	911	2,453	6,481	586,052	9.89	10	1,048	2,919	8,231	1,124,135	12.13
3rd Quintile	4,364	15	822	2,323	6,197	1,053,229	12.07	17	994	2,839	8,440	1,323,156	11.06
4th Quintile	4,276	14	1,077	2,772	7,763	313,383	4.68	14	1,285	3,558	10,977	355,652	6.06
5th Quintile	3,342	15	1,216	3,258	8,732	701,763	6.12	17	1,488	4,266	12,366	919,088	6.64
Population Density													
Large MSA[1]	8,192	15	1,418	3,593	8,988	1,053,229	8.36	17	1,717	4,648	12,779	1,323,156	8.06
Medium to Small MSA[1]	9,833	8	749	2,019	5,606	586,052	7.84	10	874	2,423	7,178	876,058	9.61
Non-MSA,[1] Urban	380	48	623	1,562	4,150	674,933	42.44	52	742	1,980	5,241	854,597	35.92
Non-MSA,[1] Rural	1,054	35	593	1,682	4,927	1,028,039	34.20	38	717	1,978	6,264	1,550,385	35.14
Group Quarters													
Group	312	15	349	872	1,924	23,423	3.63	17	429	1,061	2,898	34,734	4.01
Non-Group	19,147	8	983	2,606	7,093	1,053,229	9.72	10	1,150	3,234	9,414	1,550,385	10.09

[1] This step used demographic variables from screener data for all selected person pairs; DU = dwelling unit, MSA = metropolitan statistical area, NR = nonresponse adjustment, PR = pair, Res = respondent, SDU = screener dwelling unit.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.

Table L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev)

		Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
Domain	***n***	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**	**Min**	**Q1[2]**	**Med**	**Q3[2]**	**Max**	**UWE[3]**
Total	19,459	10	1,129	3,171	9,270	1,550,385	10.18	4	1,082	3,123	9,214	1,234,015	9.58	3	1,068	3,112	9,208	1,237,337	9.36
Pair Age Group																			
12-17, 12-17	3,668	10	593	1,177	2,391	34,664	2.58	4	542	1,168	2,393	30,611	2.50	3	534	1,162	2,395	32,280	2.52
12-17, 18-25	2,759	15	847	1,791	3,683	37,794	2.42	6	815	1,821	3,816	25,541	2.31	6	813	1,829	3,856	24,364	2.31
12-17, 26-34	658	143	2,030	4,412	8,943	151,613	3.25	89	1,938	4,295	9,294	102,559	3.01	87	1,884	4,201	9,418	112,015	3.09
12-17, 35-49	2,812	176	2,886	5,847	12,501	148,919	2.69	130	2,567	5,540	12,475	142,220	2.83	105	2,514	5,526	12,426	143,455	2.82
12-17, 50+	631	258	5,228	10,765	22,907	230,684	2.72	184	4,418	10,190	23,400	196,486	2.77	179	4,462	10,151	23,711	192,088	2.71
18-25, 18-25	3,901	17	644	1,524	3,865	44,826	2.83	15	580	1,562	4,185	44,635	2.77	14	568	1,556	4,193	32,458	2.71
18-25, 26-34	794	52	1,764	3,722	10,598	323,710	4.78	40	1,516	3,814	9,782	153,541	3.94	36	1,448	3,764	9,723	151,757	3.98
18-25, 35-49	1,053	183	3,389	8,145	21,593	184,727	2.82	172	3,071	7,900	21,308	177,102	2.80	158	3,082	7,802	21,369	154,716	2.75
18-25, 50+	660	289	7,649	15,463	33,878	296,605	2.59	432	6,714	15,392	33,732	299,472	2.64	402	6,780	15,470	34,493	279,736	2.57
26-34, 26-34	604	207	3,971	7,400	14,503	481,267	6.58	232	3,844	8,467	16,566	727,153	7.05	216	3,645	8,417	16,631	624,327	6.46
26-34, 35-49	320	412	6,938	12,531	27,491	776,131	5.14	137	5,215	12,267	24,629	1,004,156	7.79	130	5,035	11,784	24,653	1,011,952	7.98
26-34, 50+	177	991	12,555	30,990	77,130	919,088	3.28	812	12,545	30,285	91,722	633,208	2.64	756	12,324	30,567	91,602	549,404	2.55
35-49, 35-49	487	463	7,217	14,279	29,236	1,550,385	10.96	164	6,271	15,001	32,702	1,234,015	8.51	155	6,301	15,391	32,738	1,237,337	8.28
35-49, 50+	272	1,038	12,849	27,776	71,681	1,024,681	3.21	671	10,431	27,472	74,724	760,439	2.95	620	10,194	26,379	74,748	612,413	2.84
50+, 50+	663	1,612	20,423	36,884	70,196	876,058	2.67	1,647	18,150	35,352	69,493	818,783	2.72	1,632	18,609	35,696	71,197	815,689	2.68
Pair Race/Ethnicity																			
Hispanic or Latino	3,134	25	1,251	3,295	10,003	1,323,156	13.84	15	1,218	3,280	9,992	1,234,015	13.21	15	1,201	3,234	9,988	1,237,337	12.94
Black or African American	2,139	21	1,286	3,663	9,282	876,058	9.71	15	1,253	3,637	9,153	818,783	9.44	15	1,247	3,647	9,107	815,689	9.47
White	11,006	29	1,102	3,054	8,995	1,550,385	10.05	23	1,093	3,116	9,136	1,010,083	8.72	21	1,078	3,108	9,117	1,011,952	8.50
Other	1,116	14	958	3,017	10,103	337,536	6.66	4	914	3,081	10,892	374,107	7.06	3	926	3,165	10,780	346,839	6.88
White & Black or African American	167	86	1,809	6,079	14,148	77,753	2.61	124	1,826	7,641	16,843	100,462	2.57	123	1,810	7,636	14,991	99,568	2.59
White & Hispanic or Latino	721	10	1,065	2,958	9,380	481,267	8.22	6	856	2,589	9,030	727,153	11.44	6	822	2,527	9,100	624,327	10.20
White & Other	715	16	881	2,689	7,897	240,232	6.02	8	608	1,904	6,289	347,902	8.03	8	568	1,854	6,261	370,874	8.47
Black or African American & Hispanic or Latino	109	98	1,474	4,559	9,426	140,505	4.00	166	1,473	4,575	10,998	165,865	4.66	145	1,364	4,448	10,885	180,929	4.82
Black or African American & Other	198	17	1,033	3,397	7,930	181,458	6.32	5	392	1,408	3,562	91,049	5.96	4	349	1,388	3,582	90,913	5.98
Hispanic or Latino & Other	154	55	1,030	3,551	9,841	206,329	6.36	32	539	2,016	6,085	109,860	6.18	30	521	1,973	6,135	109,119	6.22

Table L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev) (continued)

Domain	n	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Pair Gender																			
Male, Male	4,162	10	1,127	2,892	8,027	488,930	7.41	6	1,078	2,843	8,046	501,086	7.64	6	1,057	2,848	8,133	502,540	7.60
Female, Female	4,430	17	1,023	2,852	7,945	285,780	6.13	13	965	2,752	7,804	374,107	6.57	12	961	2,741	7,772	346,839	6.53
Male, Female	10,867	14	1,170	3,460	10,465	1,550,385	11.13	4	1,141	3,386	10,381	1,234,015	10.20	3	1,121	3,357	10,468	1,237,337	9.92
Household Size																			
Two	4,660	17	856	2,654	11,109	292,988	5.19	16	774	2,609	11,083	280,634	5.16	14	743	2,532	11,015	288,931	5.26
Three	5,248	10	1,189	3,534	8,651	1,550,385	12.87	5	1,150	3,349	8,560	1,010,083	10.50	4	1,158	3,356	8,647	1,002,962	10.00
Four or More	9,551	14	1,264	3,187	8,995	1,323,156	11.36	4	1,226	3,180	8,946	1,234,015	11.35	3	1,224	3,186	8,955	1,237,337	11.11
Census Region																			
Northeast	3,856	10	813	2,697	8,171	1,124,135	10.75	4	754	2,559	8,056	1,010,083	11.36	3	746	2,522	8,063	1,011,952	11.29
South	5,867	17	1,403	4,051	11,527	1,323,156	8.68	5	1,362	3,928	11,269	1,234,015	8.81	4	1,364	3,931	11,279	1,237,337	8.81
Midwest	5,455	30	1,083	2,645	6,808	1,550,385	14.22	18	1,043	2,601	6,923	804,009	9.01	16	1,023	2,563	6,817	639,872	8.34
West	4,281	36	1,161	3,392	10,802	1,024,681	8.68	22	1,132	3,385	10,721	760,439	8.69	20	1,105	3,351	10,734	624,327	8.33
Quarter																			
Quarter1	4,613	14	1,351	3,582	10,437	854,597	7.22	4	1,274	3,548	10,107	805,478	7.47	3	1,247	3,494	10,168	806,173	7.45
Quarter2	5,115	25	1,071	3,131	8,922	417,070	6.38	9	1,021	3,015	8,778	506,343	6.85	8	1,001	3,000	8,776	515,372	6.86
Quarter3	5,093	10	963	2,769	8,219	1,024,681	12.79	6	917	2,717	8,257	884,682	12.27	6	910	2,724	8,293	890,858	11.63
Quarter4	4,638	15	1,189	3,328	9,797	1,550,385	14.07	6	1,170	3,336	9,836	1,234,015	11.60	6	1,151	3,312	9,789	1,237,337	11.37
% Hispanic or Latino in Segment																			
50-100%	1,537	42	1,594	4,086	11,989	854,597	9.02	34	1,508	4,019	12,166	884,682	9.65	33	1,482	3,980	12,144	890,858	9.73
10-<50%	3,726	35	1,482	4,402	11,789	1,323,156	10.37	9	1,434	4,263	11,836	1,234,015	9.54	8	1,399	4,222	11,854	1,237,337	9.27
<10%	14,196	10	1,024	2,844	8,288	1,550,385	9.90	4	968	2,790	8,241	1,010,083	9.16	3	960	2,779	8,206	1,011,952	8.90
% Black or African American in Segment																			
50-100%	1,482	17	1,232	3,602	10,074	488,930	6.27	5	1,175	3,589	10,217	501,086	6.35	4	1,165	3,573	10,398	502,540	6.32
10-<50%	2,974	29	1,287	3,799	10,451	1,024,681	9.57	18	1,257	3,738	10,228	805,478	8.61	16	1,267	3,750	10,205	806,173	8.21
<10%	15,003	10	1,089	3,042	8,915	1,550,385	10.70	4	1,034	2,978	8,935	1,234,015	10.13	3	1,018	2,949	8,924	1,237,337	9.94
% Owner-Occupied DUs[1] in Segment																			
50-100%	14,866	10	1,248	3,505	10,121	1,550,385	9.97	4	1,216	3,410	9,934	1,234,015	9.37	3	1,209	3,401	9,941	1,237,337	9.12
10-<50%	3,600	17	991	2,731	8,123	1,124,135	9.53	9	929	2,762	8,444	1,010,083	9.02	8	909	2,733	8,434	1,002,962	9.05
<10%	993	17	458	1,220	3,258	133,448	6.99	5	364	1,097	3,271	123,637	7.55	4	357	1,098	3,192	117,144	7.34
Combined Median Rent/Housing Value																			
1st Quintile	3,227	15	955	2,526	6,979	1,550,385	21.52	6	943	2,599	7,106	1,004,156	15.44	6	933	2,629	7,213	1,011,952	14.03
2nd Quintile	4,250	10	1,048	2,919	8,231	1,124,135	12.13	6	1,020	2,887	8,310	1,010,083	12.22	6	1,008	2,860	8,337	1,002,962	12.24
3rd Quintile	4,364	17	994	2,839	8,440	1,323,156	11.06	5	944	2,762	8,242	1,234,015	11.13	4	921	2,724	8,185	1,237,337	11.29
4th Quintile	4,276	14	1,285	3,558	10,977	355,652	6.06	4	1,220	3,461	10,809	399,094	6.23	3	1,209	3,424	10,736	413,094	6.20
5th Quintile	3,342	17	1,488	4,266	12,366	919,088	6.64	18	1,409	4,103	12,186	727,153	6.73	16	1,419	4,111	12,119	624,327	6.41

Table L.3 2012 NSDUH Respondent Pair-Level Weight Summary Statistics (res.pr.ps and res.pr.ev) (continued)

Domain	*n*	Before res.pr.ps[1] (SDUWT*PRWT10*...*PRWT12)						After res.pr.ps[1] (SDUWT*PRWT10*...*PRWT13)						Final Weight: After res.pr.ev[1] (SDUWT*PRWT10*...*PRWT14)					
		Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]	Min	Q1[2]	Med	Q3[2]	Max	UWE[3]
Population Density																			
Large MSA[1]	8,192	17	1,717	4,648	12,779	1,323,156	8.06	5	1,708	4,620	12,943	1,234,015	7.56	4	1,692	4,589	12,922	1,237,337	7.40
Medium to Small MSA[1]	9,833	10	874	2,423	7,178	876,058	9.61	4	807	2,307	6,911	1,004,156	11.13	3	788	2,290	6,933	1,011,952	11.08
Non-MSA,[1] Urban	380	52	742	1,980	5,241	854,597	35.92	16	762	1,878	5,022	805,478	32.27	16	737	1,852	5,022	806,173	32.41
Non-MSA,[1] Rural	1,054	38	717	1,978	6,264	1,550,385	35.14	30	852	2,366	6,703	804,009	13.38	29	859	2,382	6,979	639,872	10.49
Group Quarters																			
Group	312	17	429	1,061	2,898	34,734	4.01	25	408	910	2,667	35,420	3.79	25	418	877	2,730	28,121	3.70
Non-Group	19,147	10	1,150	3,234	9,414	1,550,385	10.09	4	1,109	3,185	9,334	1,234,015	9.49	3	1,093	3,175	9,352	1,237,337	9.28
Pair Relationship Domain[4]																			
Parent-Child (12-14)	1,981	83	2,456	5,444	11,989	163,632	2.95	80	2,413	5,537	12,561	196,486	3.13	75	2,355	5,499	12,516	192,088	3.14
Parent-Child (12-17)	3,748	15	2,755	5,933	12,968	230,684	2.95	6	2,630	5,869	13,546	196,486	3.04	6	2,547	5,798	13,561	192,088	3.04
Parent-Child (12-20)	4,443	15	2,915	6,347	14,088	296,605	3.18	6	2,825	6,450	14,879	299,472	3.27	6	2,793	6,413	14,820	279,736	3.21
Sibling (12-14)-Sibling (15-17)	2,191	14	588	1,175	2,404	34,664	2.58	4	562	1,193	2,423	30,611	2.44	3	555	1,184	2,449	32,280	2.47
Sibling (12-17)-Sibling (18-25)	2,477	16	844	1,789	3,667	37,794	2.36	8	828	1,843	3,811	23,928	2.28	8	823	1,839	3,849	24,220	2.28
Spouse-Spouse/ Partner-Partner	3,664	30	1,097	4,481	16,693	1,550,385	10.77	16	1,108	4,420	17,392	1,234,015	9.77	14	1,065	4,375	17,434	1,237,337	9.62
Spouse-Spouse/ Partner-Partner with Children (Younger Than 18)	1,640	51	1,055	3,442	10,817	1,323,156	16.13	70	1,595	5,143	15,628	1,234,015	12.24	69	1,575	5,108	15,656	1,237,337	12.17

[1] This step used demographic variables from questionnaire data for all selected person pairs; DU = dwelling unit, EV = extreme value adjustment, MSA = metropolitan statistical area, PR = pair, PS = poststratification adjustment, Res = respondent, SDU = screener dwelling unit.
[2] Q1 and Q3 refer to the first and third quartile of the weight distribution.
[3] Unequal weighting effect (UWE) is defined as $1 + [(n - 1)/n]*CV^2$, where CV = coefficient of variation of weights.
[4] Parent-child (15-17) was not included here since extreme values were not controlled with this domain.

Appendix M: Pair Analysis Manual Excerpt

Appendix M: Pair Analysis Manual Excerpt

This appendix provides background information on pair data analysis for the National Survey on Drug Use and Health (NSDUH). This excerpt is from Section 3.2 in *How To Prepare and Analyze Pair Data in the National Survey on Drug Use and Health* by Frechtel, Warren, and Porter (in press).

Inferential Population and Multiplicities

There are different perspectives through which pair data can be analyzed: (1) with pairs as the focus, or (2) with one member of the pair as the focus. When the focus is on the pair, the *PRANALWT* variable can be used to weight the data directly with no adjustments. However, when the focus is on one member of the pair, an adjustment often needs to be made to the weight to account for this. For example, the analysis outlined in Table 1 is focused on the child's behavior when the father talked to him or her about substance use. However, if the analyst was interested in the child's behavior when a parent talked to him or her about substance use, regardless of the gender of the parent, the weight would need to be adjusted to account for the fact that the child may be a member of more than one parent-child pair. For the analysis in Table 1, there is no multiple counting problem as long as an assumption is made that no child lives with more than one father.

Adjustments for the multiple counting problem are done using "multiplicities" (Chromy & Singh, 2001). These multiplicities have been computed for the analyst and can be accounted for by simply adjusting the weight variable (described in Section 4.4). Analysts should exercise care or seek assistance in computing multiplicities for any pair types not listed. The process by which the multiplicities are created is described in detail in Chapter 10 of Frechtel et al. (2013). Table 5 lists the pair domains for which multiplicities can be used to perform person-level analyses.

Table 5. Pair Domains and Multiplicities

Pair Domain			Multiplicity Variable	
***IRPRREL* Levels**	**Description**	**Focus[1]**	**Name**	**Description**
1	Parent-child, child aged 12-14	Parent	*IRMPCP14*	Number of children aged 12-14 living with responding parent
		Child	*IRMPCC14*	Number of parents living with responding child aged 12-14
1,2	Parent-child, child aged 12-17	Parent	*IRMPCP17*	Number of children aged 12-17 living with responding parent
		Child	*IRMPCC17*	Number of parents living with responding child aged 12-17
1,2,3	Parent-child, child aged 12-20	Parent	*IRMPCP20*	Number of children aged 12-20 living with responding parent
		Child	*IRMPCC20*	Number of parents living with responding child aged 12-20
2	Parent-child, child aged 15-17	Parent	*IRMPCP57*	Number of children aged 15-17 living with responding parent
		Child	*IRMPCC57*	Number of parents living with responding child aged 15-17
5	Sibling-sibling, older sibling aged 15-17, younger sibling aged 12-14	Older sibling	*IRMS1417*	Number of siblings aged 12-14 living with responding sibling aged 15-17
		Younger sibling	*IRMS1714*	Number of siblings aged 15-17 living with responding sibling aged 12-14
6	Sibling-sibling, older sibling aged 18-25, younger sibling aged 12-17	Older sibling	*IRMS1725*	Number of siblings aged 12-17 living with responding sibling aged 18-25
		Younger sibling	*IRMS2517*	Number of siblings aged 18-25 living with responding sibling aged 12-17
8,9	Spouse-spouse and partner-partner	No multiplicity necessary: assume only one spouse per person		
8	Spouse-spouse and partner-partner, with children aged 0-17	No multiplicity necessary: assume only one spouse per person		

[1] No weight adjustment is needed when the inferential focus is on the pair.

To help clarify this concept, some bulleted examples are included below. A full list of multiplicities and definitions is included in Table 5 above. The examples are based on the analysis in Table 6, which is the same as the analysis in Table 1 except that the parent can be a father or a mother.

Table 6. Example of Pair-Level Analysis Requiring Multiplicities

Parent Reports Talking to Child Aged 12-17 about Substance Use	Child Used an Illicit Drug in the Past Year	
	Yes	**No**
Yes	(1,1)	(1,2)
No	(2,1)	(2,2)

- To populate Table 6 with estimates related to children (i.e., row percentages), the proper weight is *PRANALWT/IRMPCC17*. For example, this weight could be used to estimate the probability that a child aged 12 to 17 used an illicit drug in the past year, given that a parent talked to him or her about substance use in the past year (cell (1,1) in the table).
- If Table 6 instead showed estimates related to parents (i.e., column percentages), the proper weight would be *PRANALWT/IRMPCP17*. For example, this weight could be used to estimate the probability that a parent talked to his or her child aged 12 to 17 about substance use in the past year, given that his or her child used an illicit drug in the past year (again, cell (1,1) in the table).
- As stated above, if estimates related to children are desired, but interest is restricted to either fathers or mothers, standard practice is to assume the presence of no more than one father and no more than one mother in the dwelling unit. In these cases, no multiplicity is necessary, and *PRANALWT* can be used as the analysis weight. For example, *PRANALWT* should be used to estimate the probability that a child aged 12 to 17 used an illicit drug in the past year, given that his or her mother talked to him or her about substance use in the past year.
- If Table 6 showed estimates related to parent-child pairs and did not focus on either member of the pair (i.e., cell percentages), then no multiplicity would be necessary. The variable *PRANALWT* is designed for this. An analysis like this is of questionable value because of the multiple counting problem, however. The inferential population includes all parent-child pairs where the child is aged 12 to 17. Many persons are represented in more than one pair. A dwelling unit with two parents and three children aged 12 to 17 would represent six pairs in the population, with each parent being a member of three of the six and each child being a member of two of the six. A dwelling unit with one parent and one child aged 12 to 17 would represent only one pair in the population. Analyses that focus on only one member of the pair usually have a more natural interpretation.

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

STATISTICAL INFERENCE REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.108.001.002
Deliverable No. 41

Authors:

Jeremy Aldworth
James R. Chromy
Teresa R. Davis
Misty S. Foster
Lisa E. Packer
Kathryn Spagnola
Lauren Warren

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

December 2013

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

STATISTICAL INFERENCE REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.108.001.002
Deliverable No. 41

Authors:

Jeremy Aldworth
James R. Chromy
Teresa R. Davis
Misty S. Foster
Lisa E. Packer
Kathryn Spagnola
Lauren Warren

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

December 2013

Acknowledgments

This publication was developed for the Substance Abuse and Mental Health Services Administration (SAMHSA), Center for Behavioral Health Statistics and Quality (CBHSQ), by RTI International (a trade name of Research Triangle Institute), Research Triangle Park, North Carolina, under Contract No. HHSS283201000003C. Significant contributors at RTI include Jeremy Aldworth, James R. Chromy, Teresa R. Davis, Misty S. Foster, Lisa E. Packer, Kathryn Spagnola, and Lauren Warren. Valerie Garner formatted the report, and Claudia Clark and Margaret Smith copyedited the report. Thomas G. Virag is the RTI Project Director.

Table of Contents

Section

1. Introduction 1

2. Background 3

3. Prevalence Rates 7
   3.1 Mental Illness 7
   3.2 Adult Major Depressive Episode (MDE) 12
   3.3 Serious Psychological Distress (SPD) 13
   3.4 Decennial Census Effects on NSDUH Substance Use and Mental Health Estimates 13
   3.5 Revised Estimates for 2006 to 2010 14

4. Missingness 17
   4.1 Potential Estimation Bias Due to Missingness 17
   4.2 Variance Estimation in the Presence of Missingness 18

5. Sampling Error 21

6. Degrees of Freedom 25

7. Statistical Significance of Differences 27

8. Confidence Intervals 29

9. Incidence Estimates 31

10. Suppression of Estimates with Low Precision 35

References 39

Appendix A: Documentation for Conducting Various Statistical Procedures: SAS and SUDAAN Examples A-1

1. Introduction

Statistical inference occurs whenever data obtained from sample observations belonging to and considered representative of a larger target population are used to make generalizations concerning the larger population. The target population for the 2012 National Survey on Drug Use and Health (NSDUH)[1] was the U.S. civilian, noninstitutionalized population aged 12 or older (at the time of their interview) in 2012. Measurements for this target population were the responses to the survey questions provided by persons participating in the 2012 survey.

Statistical inferences concerning characteristics of interest for this population and various subpopulations are presented in the form of estimates derived from the sample data collected. Examples of the inferences made from the 2012 NSDUH data are presented in the 2012 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2013a) and the 2012 summary of national findings report (CBHSQ, 2013d) and include estimates of the number of persons who were substance users during the past month, past year, and their lifetime, as well as the associated percentages (prevalence rates) of substance use for these reference periods. Inferences also were made for such categories as substance initiation; risk and protective factors; substance dependence, dependence or abuse, and treatment. Estimates of measures related to mental health problems are presented in the 2012 mental health detailed tables (CBHSQ, 2013b) and the 2012 mental health findings report (CBHSQ, 2013c).

The focus of this report is to describe the statistical inference procedures used to produce design-based estimates as presented in the 2012 detailed tables, the 2012 mental health detailed tables, the 2012 national findings report, and the 2012 mental health findings report.[2] The statistical procedures and information found in this report can also be generally applied to analyses based on the public use file as well as the restricted-use file available through the data portal.[3] This report is organized as follows: Section 2 provides background information concerning the 2012 NSDUH; Section 3 discusses the prevalence rates and how they were calculated; Section 4 briefly discusses how missing item responses of variables that are not imputed may lead to biased estimates; Section 5 discusses sampling errors and how they were calculated; Section 6 describes the degrees of freedom that were used when comparing estimates; and Section 7 discusses how the statistical significance of differences between estimates was determined. Section 8 discusses confidence interval estimation, and Section 9 describes how past year incidence of drug use was computed. Finally, Section 10 discusses the conditions under which estimates with low precision were suppressed. Appendix A contains examples that demonstrate how to conduct various statistical procedures documented within this report using SAS® and SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2008).

---

[1] Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

[2] Users of the 2012 public use file (CBHSQ/SAMHSA, 2013) may find inconsistencies in the variable names referenced in this report, the information presented in Table 3 in Section 5, and other specific numbers presented in this report (i.e., degrees of freedom). The specific information referenced in this report is based on the restricted-use data set that was used to create the 2012 detailed tables (CBHSQ, 2013a), the 2012 mental health detailed tables (CBHSQ, 2013b), the 2012 mental health findings report (CBHSQ, 2013c), and the 2012 national findings report (CBHSQ, 2013d).

[3] The data portal can be found at http://www.icpsr.umich.edu/icpsrweb/content/SAMHDA/dataportal.html.

2. Background

The 2012 National Survey on Drug Use and Health (NSDUH) is an extension of a coordinated 5-year sample design providing estimates for all 50 States and the District of Columbia for the years 2005 through 2009, and then continuing through 2012. The survey is conducted using computer-assisted interviewing methods for the screening and interviewing of selected respondents. The respondent universe is the civilian, noninstitutionalized population aged 12 or older residing within the United States and the District of Columbia. Persons excluded from the universe include active-duty military personnel, persons with no fixed household address (e.g., homeless and/or transient persons not in shelters), and residents of institutional group quarters, such as correctional facilities, nursing homes, mental institutions, and long-term hospitals.

The coordinated design for 2005 through 2009 facilitated a 50 percent overlap in second-stage units (area segments) within each successive 2-year period from 2005 through 2009. The 2010, 2011, and 2012 NSDUHs continued the 50 percent overlap by retaining half of the second-stage units from the previous survey year. Those segments not retained from the previous year are considered "retired" from use; that is, these segments will not be used to field another main study sample.

Because the coordinated design enables estimates to be developed by State in all 50 States and the District of Columbia, States may be viewed as the first level of stratification as well as a reporting variable. Eight States were designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) with target sample sizes of 3,600. In 2012, sample sizes in these States ranged from 3,544 to 3,687 respondents. For the remaining 42 States and the District of Columbia, the target size was 900. Sample sizes in these States ranged from 829 to 976. State estimates combining multiple years of data and using either small area estimation[4] or direct estimation have been tabulated.

States were first stratified into a total of 900 State sampling regions (SSRs) (48 regions in each large sample State and 12 regions in each small sample State). These regions were contiguous geographic areas designed to yield on average the same number of interviews.[5] Unlike the 1999–2001 National Household Surveys on Drug Abuse and the 2002–2004 NSDUHs in which the first-stage sampling units were clusters of census blocks called area

[4] Small area estimation is a hierarchical Bayes modeling technique used to produce State-level estimates for a selected number of measures. For more details, see the *State Estimates of Substance Use and Mental Disorders from the 2009-2010 National Surveys on Drug Use and Health* (Hughes, Muhuri, Sathe, & Spagnola, 2012).

[5] Areas were defined using 2000 census geography. Dwelling units and population counts were obtained from the 2000 census data supplemented with revised population counts from Nielsen Claritas (see http://www.nielsen.com/us/en.html).

segments, the first stage of selection for the 2005–2012 surveys was census tracts.[6] This stage was included to contain sample segments within a single census tract to the extent possible.[7]

A total of 48 census tracts per SSR were selected, and within these sampled census tracts, adjacent census blocks were combined to form the second-stage sampling units or area segments. Although only 24 segments were needed to support the coordinated 5-year sample, an additional 24 segments were selected to support any supplemental studies that the Substance Abuse and Mental Health Services Administration may choose to field. These 24 segments constitute the reserve sample and were available for use in 2010, 2011, and 2012. Eight reserve sample segments per SSR were fielded during the 2012 survey year. Four of these segments were retained from the 2011 survey, and four were selected for use in the 2012 survey. These sampled segments were allocated equally into four separate samples, one for each 3-month period (calendar quarter) during the year. That is, a sample of addresses was selected from two segments in each calendar quarter so that the survey was essentially continuous in the field.

The overall design remained the same beginning with the 2002 NSDUH and continuing through the 2012 NSDUH. Survey respondents were given a $30 incentive payment for participation. Also, a pair-sampling strategy was implemented that increased the number of pairs selected in dwelling units with older persons on the roster (Chromy & Penne, 2002).

As was done in the 2008–2011 NSDUHs, a Mental Health Surveillance Study (MHSS) was embedded in the 2012 NSDUH. Each respondent in a subsample of adults (about 1,500 in 2008, 500 in 2009 and 2010, and 1,500 in 2011 and 2012) who had completed the NSDUH interview was administered the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) (First, Spitzer, Gibbon, & Williams, 2002).[8] The SCID was adapted for this study and was administered via paper and pencil over the telephone approximately 2 to 4 weeks after the NSDUH interview. In 2008, a split-sample MHSS was conducted to develop models using the SCID data that would use the Kessler-6 nonspecific psychological distress scale and two competing functional impairment scales to generate prevalence estimates of serious mental illness among adults aged 18 or older for the entire sample. Based on the results from the 2008 MHSS, a modified World Health Organization Disability Assessment Schedule (Rehm et al., 1999) was adopted for the 2009–2012 surveys. Using data from the 2008–2012 MHSS, an improved mental health prediction model was developed. The revised and improved model was used to predict mental illness starting with the 2012 NSDUH (more details on this model can be found in Section 3.1 of this report). For more information about the MHSS sample design, see the Sample Design Report in the *2012 NSDUH Methodological Resource Book* (Morton et al., 2013).

---

[6] Census tracts are relatively permanent statistical subdivisions of counties and provide a stable set of geographic units across decennial census periods.

[7] Some census tracts had to be aggregated in order to meet the minimum dwelling unit (DU) requirement of 150 DUs in urban areas and 100 DUs in rural areas.

[8] "DSM-IV-TR" stands for the *Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Text Revision* (American Psychiatric Association, 2008).

The final respondent sample of 68,309 persons for the 2012 NSDUH provides a sufficient sample to create domain estimates for a broad range of ages and other demographic categories. Individual observations are weighted so that the weighted sample represents the civilian, noninstitutionalized population aged 12 or older for the general U.S. population and for each of the individual States. The person-level weights in NSDUH are calibrated to population estimates (or control totals) obtained from the U.S. Census Bureau. For more information on the sampling weight calibration in the 2012 NSDUH, see Chen et al. (2014).

3. Prevalence Rates

The national prevalence rates were computed using a multiprocedure package called SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2008). The final, nonresponse-adjusted, and poststratified analysis weights were used in SUDAAN to compute unbiased design-based drug use estimates. Appendix A contains an example that demonstrates how to compute the prevalence rates using SUDAAN procedures as defined below.

Prevalence rates are the proportions of the population who exhibit characteristics of interest (such as substance use). Let $\hat{p}_d$ represent the prevalence rate of interest for domain *d*. Then $\hat{p}_d$ would be defined as the ratio

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

where $\hat{Y}_d = \sum_{i \in S} w_i \delta_i y_i$ represents the estimated number of persons exhibiting the characteristic of interest in domain *d*, $\hat{N}_d = \sum_{i \in S} w_i \delta_i$ represents the estimated population total for domain *d*, *S* represents the sample, w_i represents the analysis weight, δ_i represents an indicator variable that is defined as 1 if the *i*th sample unit is in domain *d* and is equal to 0 otherwise, and y_i represents an indicator variable that is defined as 1 if the *i*th sample unit exhibits the characteristic of interest and is equal to 0 otherwise.

For certain populations of interest, sample sizes may not be adequate to support inferences using only 1 year of survey data. In these cases, estimates were produced from annual averages based on combined data from 2 or more survey years, and they are clearly labeled in the detailed tables. The data were combined for the 2009–2010, 2011–2012, and 2009–2012 surveys to obtain annual averages, and then the prevalence rates were computed in SUDAAN as described above. The annual averages were derived by concatenating the data for the respective years and dividing the analysis weights by a factor that varied depending on the number of years of concatenated data. The weight was divided by a factor of 2 for 2 years of concatenated data and a factor of 4 for 4 years of concatenated data.

3.1 Mental Illness

The Substance Abuse and Mental Health Services Administration (SAMHSA) has been publishing estimates of the prevalence of past year serious mental illness (SMI) and any mental illness (AMI) among adults aged 18 or older since the release of the 2008 National Survey on Drug Use and Health (NSDUH) national findings report (Office of Applied Studies [OAS], 2009b). Estimates were based on a prediction model for mental illness developed using the 2008 data from the Mental Health Surveillance Study (MHSS) (referred to as the 2008 World Health Organization Disability Assessment Schedule [WHODAS] model). The 2008 NSDUH included a split sample, in which half the respondents (approximately 750) were administered the

WHODAS and the other half the Sheehan Disability Scale (SDS). Two models were used to predict SMI for 2008, one for each impairment scale (WHODAS and SDS). The 2008 models for SMI were chosen so that estimates from the WHODAS and SDS samples were approximately equal; hence, SMI estimates for 2008 were based on both samples. The WHODAS model was determined to be a better predictor of SMI than the SDS model; therefore, starting in 2009, only the WHODAS impairment scale was administered in the NSDUH and used for estimating all levels of mental illness (SMI, mental illness [MI], AMI, mild mental illness [LMI], moderate mental illness [MMI], and serious or moderate mental illness [SMMI]).

Although SAMHSA continued to obtain clinical interviews after 2008, estimates of mental illness from the 2009, 2010, and 2011 NSDUHs have been based on the WHODAS model developed from the 2008 clinical assessment sample. The same model was applied to each year's NSDUH data to provide consistency in mental illness comparisons across the years. Producing a new model each year based on the small annual clinical samples (only 500 interviews in 2009 and 2010) would have resulted in large changes in the model parameters and corresponding prevalence rates due to sampling error, making it impossible to detect real trends in mental illness over time. Furthermore, an evaluation of the 2008 model, using the 2009 NSDUH clinical data, found that the model could not be significantly improved with the additional 500-case 2009 clinical sample. The clinical follow-up study, which started in 2008 and continued until 2012, led to a nationally representative sample of approximately 5,000 cases assigned to the WHODAS questions that were used to develop an improved mental illness prediction model (referred to as the 2012 WHODAS model). This revised and improved model was used to predict mental illness starting with the 2012 NSDUH and incorporates the NSDUH respondent's age and indicators of past year suicidal thoughts and depression, along with the variables that were specified in the 2008 model (e.g., variables for the Kessler-6 [K6] scale and the WHODAS), leading to more accurate estimates of mental illness (see below for details on the 2012 model and revised methodology).

For the 2012 mental health detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2013b), both the 2012 estimates and the previous 2008–2011 estimates were based on the revised model. As of October 2013, the 2008 detailed tables (CBHSQ, 2009) and the 2009–2011 mental health detailed tables (CBHSQ, 2010; 2012a; 2012c) containing estimates for past year mental illness for adults have been revised based on the 2012 model. It is recommended that the mental illness variables derived from the 2012 model should not be used when analyzing variables for past year suicidal thoughts, past year major depressive episode (MDE), the K6 scale, or the WHODAS, and it is also recommended that the mental illness variables derived from the 2012 model not be used when analyzing other closely linked variables (including past year suicide attempts, past year suicide plans, medical treatment for suicide attempts, lifetime MDE, serious psychological distress [SPD], or components used in the K6 or WHODAS scales). For detailed information on model revisions to the mental illness items, see Section B.4.3 in Appendix B of the *Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings*. As with the mental illness estimates based on the 2008 model, the mental illness estimates based on the 2012 model are not comparable with SMI estimates produced from NSDUH data before 2004, and SMI estimates were not produced from 2004 to 2007; thus, long-term trend estimates are not available for SMI.

2012 SMI Prediction Model

The 2012 model is a prediction model for mental illness. This prediction model is a weighted logistic regression. The response variable Y was defined so that $Y = 1$ when an SMI diagnosis was positive based on the clinical interview; otherwise, $Y = 0$. If $\mathbf{x}$ is a vector of realized explanatory variables, then the response probability $\pi = \Pr(Y = 1 \mid \mathbf{x})$ can be estimated using a weighted logistic regression model. For further technical details on the 2012 prediction models and the impact of the revised model on the 2008–2011 estimates, see the *2012 Mental Health Surveillance Study: Design and Estimation Report* (CBHSQ, in press) or Section B.4.3 in Appendix B of the *Results from the 2012 National Survey on Drug Use and Health: Mental Health Findings* (CBHSQ, 2013c).

The 2012 SMI prediction model was fit with data from 4,912 WHODAS MHSS respondents from 2008 through 2012, excluding one case from 2008 and one case from 2009 that were dropped because of data errors. The final WHODAS calibration model for the 2012 prediction model for SMI was determined as

$$\begin{aligned}\text{logit}(\hat{\pi}) = \log[\hat{\pi} / (1 - \hat{\pi})] = -5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s \\ + 1.1267330\,X_m + 0.1059137X_a \end{aligned} \quad (1)$$

or

$$\hat{\pi} = \frac{1}{1 + \exp[-(-5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s + 1.1267330\,X_m + 0.1059137X_a)}},$$

where $\hat{\pi}$ refers to the estimate of the SMI response probability π.

These covariates in equation (1) came from the main NSDUH interview data:

X_k = *Alternative Past Year K6 Score*: Past year K6 score of less than 8 recoded as 0; past year K6 score of 8 to 24 recoded as 1 to 17.

X_w = *Alternative WHODAS Score*: WHODAS item score of less than 2 recoded as 0; WHODAS item score of 2 to 3 recoded as 1, then summed for a score ranging from 0 to 8.

X_s = *Serious Thoughts of Suicide in the Past Year*: Coded as 1 if "yes"; coded as 0 otherwise.

X_m = *Past Year MDE*: Coded as 1 if the criteria for past year MDE were met;[9] coded as 0 otherwise.

X_a = *Recoded Age*: Coded as age minus 18 if aged 18 to 30; coded as 12 otherwise.

[9] In this situation, the past year MDE measure is from the main NSDUH interview (i.e., not from the Structured Clinical Interview for DSM-IV). See section B.4.4 of the 2012 mental health findings report (CBHSQ, 2013c).

A cut point probability π_0 was determined, so that if $\hat{\pi} \geq \pi_0$ for a particular respondent, then he or she was predicted to be SMI positive; otherwise, he or she was predicted to be SMI negative. The cut points were chosen so that the weighted numbers of false positives and false negatives in the MHSS dataset were as close to equal as possible. The predicted SMI status for all adult NSDUH respondents was used to compute prevalence estimates of SMI. In the 2012 SMI WHODAS prediction model, the respondent is classified as having past year SMI if the predicted probability of SMI is greater than or equal to 0.260573529 (SMI cutoff point). See Table 1 for the model specifications. Table 2 contains the cutoff points for other mental illness levels.

Modified 2012 Model for the 2008 SDS Half Sample

As noted previously, the 2008 NSDUH data included a split sample. Similar to the 2008 model, the revised 2012 model also has an alternative model for the SDS data that was fit with data from the complete 2008–2012 MHSS clinical sample that contains 5,653 MHSS respondents, excluding 4 cases from 2008 and 1 case from 2009 that were dropped because of data errors.

The modified 2012 SMI prediction model for the SDS half sample was

$$\text{logit}(\hat{\pi}) = \log[\hat{\pi} / (1-\hat{\pi})] = -5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a \qquad (2)$$

or

$$\hat{\pi} = \frac{1}{1+\exp[-(-5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a)}.$$

All of the covariates in equation (2) also appeared in equation (1).

Similar to the WHODAS model, a cut point probability π_0 was determined, so that if $\hat{\pi} \geq \pi_0$ for a particular respondent, then he or she was predicted to be SMI positive; otherwise, he or she was predicted to be SMI negative. The cut points were chosen so that the weighted numbers of false positives and false negatives in the MHSS dataset were as close to equal as possible. In the 2012 SMI SDS half sample prediction model, the respondent is classified as having past year SMI if the predicated probability of SMI is greater than or equal to 0.236434 (SMI cutoff point). Although the SDS half sample prediction model was fit across all years and the cutoff points were determined based on all years, the cutoff points were only used for the main study respondents in the 2008 sample B to predict the SMI positives. See Table 1 for the model specifications. Table 2 contains the cutoff points for other mental illness levels.

Table 1. Final SMI Prediction Models in the 2008–2012 MHSS

	Beta	Beta SE	*T* Statistic	*P* Value	*DF*	Wald *P* Value[1]
WHODAS Sample (2008A-2012)						
Intercept	-5.9726640	0.3201	-18.6586	0.0000		
Alt PY K6	0.0873416	0.0248	3.5247	0.0009	1	0.0009
Alt WHODAS	0.3385193	0.0349	9.7034	0.0000	1	0.0000
PY Suicidal Thoughts	1.9552664	0.2164	9.0342	0.0000	1	0.0000
PY MDE	1.1267330	0.2196	5.1308	0.0000	1	0.0000
Age1830	0.1059137	0.0244	4.3380	0.0001	1	0.0001
WHODAS and SDS Samples (2008–2012)[2]						
Intercept	-5.7736246	0.3479	-16.5960	0.0000		
Alt PY K6	0.1772067	0.0190	9.3251	0.0000	1	0.0000
PY Suicidal Thoughts	1.8392433	0.1941	9.4781	0.0000	1	0.0000
PY MDE	1.6428623	0.2119	7.7528	0.0000	1	0.0000
Age1830	0.1231266	0.0259	4.7482	0.0000	1	0.0000

Age1830 = recoded age variable; Alt = alternative; *DF* = degrees of freedom; K6 = Kessler-6, a six-item psychological distress scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; PY = past year; SDS = Sheehan Disability Scale; SE = standard error; SMI = serious mental illness; WHODAS = eight-item World Health Organization Disability Assessment Schedule; 2008A = 2008 WHODAS half sample.

[1] The *p* value is obtained from the overall model fitting.

[2] The model is fit over the WHODAS and SDS samples in 2008–2012 but is used only to produce predictions for the 2008 SDS sample.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008–2012.

Table 2. Cut Point Probabilities for SMI, AMI, and SMMI, by 2012 Model

	Cut Point Probability
	WHODAS Sample (2008A-2012)
SMI	0.260573529
AMI	0.0192519810
SMMI	0.077686285365
	WHODAS and SDS Samples (2008–2012)[1]
SMI	0.236434
AMI	0.019182625
SMMI	0.06616398

AMI = any mental illness; SMI = serious mental illness; SMMI = serious or moderate mental illness; SDS = serious psychological distress; WHODAS = World Health Organization Disability Assessment Schedule.

[1] The model is fit over the WHODAS and SDS samples in 2008–2012, but the cut point predictions are only used to produce predictions for the 2008 SDS sample.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008–2012.

Weights

For the 2008 NSDUH, although SMI data for both half samples (SDS and WHODAS) could be analyzed together when using the 2008 model, the AMI, SMMI, LMI, and MMI data from the two half samples could not be combined for analysis. Under the 2012 model, both the

2008 half samples can be combined to analyze SMI and the other levels of mental illness because the 2012 models were generated so the estimates would be comparable.

Mental illness measures (i.e., SMI, AMI, SMMI, MMI, and LMI) that are defined based on the 2012 model should be analyzed using the standard analysis weight, ANALWT, for all survey years 2008 through 2012. With the revised 2012 model, both the WHODAS and SDS 2008 half samples can be combined to form single estimates and use ANALWT.

This differs from the initial recommendation for analyzing measures of mental illness besides SMI based on the 2008 model. Because of the 2008 split sample, an adjusted mental health sample weight, MHSAMPWT, was created so that the WHODAS and SDS half samples were separately representative of the civilian, noninstitutionalized population aged 18 or older. However, this weight should not be used to analyze 2008 mental illness data based on the 2012 model.

Standard Errors for Mental Illness Estimates

For the 2012 mental health findings report and the mental health detailed tables, standard errors (SEs) for mental illness estimates (SMI, AMI, SMMI, MMI, and LMI) were computed using the NSDUH dichotomous variable values without taking into account any variance introduced through using a model based on the clinical subsample data. This ignores the added error resulting from fitting the 2012 SMI model, which can be very large (CBHSQ, in press). These *conditional* SEs (conditional on the model predictions being correct) are useful when making comparisons across years and across subpopulations within years because the errors due to model fitting are nearly the same across the estimates being compared, and consequently, they roughly cancel each other out.

3.2 Adult Major Depressive Episode (MDE)

The past year adult MDE estimates shown in the 2012 mental health detailed tables (CBHSQ, 2013b) are based on the full sample as was done in the 2010 and 2011 detailed tables (CBHSQ, 2011; 2012b). This differs from the 2008 past year MDE estimates shown in both the 2008 detailed tables (OAS, 2009a) and the 2009 mental health detailed tables (CBHSQ, 2010), which were based only on the sample of adult respondents who received the WHODAS questions in the mental health questionnaire module that preceded the adult depression questionnaire module. The analysis of 2008 MDE data was restricted to only the WHODAS half sample because of apparent reporting differences (context effects) between the half sample that was administered the WHODAS and the other half sample of adult respondents who received the SDS questions (Dean & LeBaron, 2009). Both half samples have issues with context effects not seen in 2007 and previous years due to the revisions to the mental health module preceding the adult depression module. To address the break in comparability of the adult MDE data beginning in 2008 and to estimate adult MDE based on the full sample of adults from 2008, adjusted versions of lifetime and past year MDE variables for adults were created retroactively for 2005 to 2008. These variables were adjusted to make MDE estimates from the SDS half sample in 2008 and from all adult respondents for 2005 to 2007 comparable with the MDE estimates based on data from the half sample that received the WHODAS in 2008 and from all adult respondents in later years (2009 onward). The adjusted data from 2005 to 2008 can be used in conjunction

with unadjusted data from later years to estimate trends in adult MDE over the entire period from 2005 to 2012.

In the 2012 mental health detailed tables (CBHSQ, 2013b), ANALWT was used to generate all estimates of adult MDE. More information about how the statistically adjusted adult MDE variables were created can be found in Section B.4.4 in Appendix B of the 2012 mental health findings report (CBHSQ, 2013c) and in the report describing the adjustments (Aldworth, Kott, Yu, Mosquin, & Barnett-Walker, 2012).

3.3 Serious Psychological Distress (SPD)

The K6 scale, a measure of psychological distress, was used to create the SPD variable. Before 2008, the K6 consisted of one set of questions that asked adult respondents about symptoms of psychological distress in the month when they were the most depressed, anxious, or emotionally distressed in the past year. Starting in 2008, the K6 consisted of two sets of questions that asked adult respondents how frequently they experienced symptoms of psychological distress during two different periods: (1) during the past 30 days, and (2) if applicable, the month in the past year when they were at their worst emotionally. Respondents were asked about this second period only if they indicated that there was a month in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days. Because of this change, past year K6 and SPD estimates from years before 2008 were no longer comparable with estimates from 2008 onward. To address this comparability issue, adjusted versions of the past year worst K6 total score and past year SPD variables were created for each of the years from 2005 to 2007 to make the 2005–2007 past year K6 scores and past year SPD estimates comparable with their 2008–2012 counterparts.

In the 2012 mental health detailed tables (CBHSQ, 2013b), ANALWT was used to generate 2005–2012 estimates of past year SPD and 2008–2012 estimates of past month SPD. The 2012 mental health findings report (CBHSQ, 2013c) did not present SPD estimates. More information about how the adjusted K6 and SPD variables were created can be found in the report describing these adjustments (Aldworth et al., 2012).

3.4 Decennial Census Effects on NSDUH Substance Use and Mental Health Estimates

As discussed in Section 2, the person-level weights in NSDUH were calibrated to population estimates (or control totals) obtained from the U.S. Census Bureau. For the weights in 2002 through 2010, annually updated control totals based on the 2000 census were used.[10] Beginning with the 2011 weights, however, the control totals from the U.S. Census Bureau are based on the 2010 census. Two investigations were implemented at the national level to assess the effects of using control totals based on the 2010 census instead of the 2000 census. One of these investigations focused specifically on measures of substance use that are used in the 2011 national findings report (CBHSQ, 2012e) and detailed tables (CBHSQ, 2012b), while a separate

[10] In addition to the standard 2010 analysis weights poststratified to 2000 census control totals, special weights that were poststratified to 2010 census control totals are available on the 2010 NSDUH public use file (CBHSQ/SAMHSA, 2012).

analysis was conducted to evaluate the impact of the weighting changes on mental health estimates in the 2011 mental health findings report (CBHSQ, 2012d) and associated detailed tables (CBHSQ, 2012c). Because both the 2011 and 2012 NSDUH estimates are based on weights that were poststratified to population control totals that were in turn based on projections from the 2010 census, 2-year trend comparisons between 2011 and 2012 are not subject to census effects. However, cautions about trends between 2010 (or earlier years) and 2011 because of census effects also could carry over to trends between 2012 and years before 2011, especially for particular subgroups (e.g., persons reporting two or more races for both investigations, and persons reporting American Indian or Alaska Native or Native Hawaiian or Other Pacific Islander). An additional investigation was done at the State level to evaluate the impact of the model-based small area estimates (SAEs).

For more information on the impact of decennial census effects on NSDUH substance use direct estimates, see Section B.4.3 in Appendix B of the 2011 national findings report (CBHSQ, 2012e). For more information on the impact of decennial census effects on NSDUH mental health direct estimates, see Appendix A of the 2011 mental health findings report (CBHSQ, 2012d). For more information on the impact of the decennial census effects on NSDUH model-based SAEs, see http://www.samhsa.gov/data/NSDUH/2k12State/NSDUHsae2012/Index.aspx. Additionally, for more information on the sampling weight calibration in the 2011 NSDUH, see the Person-Level Sampling Weight Calibration report (Chen et al., 2013).

3.5 Revised Estimates for 2006 to 2010

During regular data collection and processing checks for the 2011 NSDUH, data errors were identified. These errors affected the data for Pennsylvania (2006–2010) and Maryland (2008–2009). Cases with erroneous data were removed from the data files, and the remaining cases were reweighted to provide representative estimates. The errors had minimal impact on the national estimates and no effect on direct estimates for the other 48 States and the District of Columbia. In reports where model-based small area estimation techniques are used, estimates for all States may be affected, even though the errors were concentrated in only two States.
In reports that do not use model-based estimates, the only estimates appreciably affected are estimates for Pennsylvania, Maryland, the mid-Atlantic division, and the Northeast region.
The 2011 detailed tables (CBHSQ, 2012b), the 2011 mental health detailed tables (CBHSQ, 2012c), the 2011 mental health findings report (CBHSQ, 2012d), and the 2011 national findings report (CBHSQ, 2012e) do not include State-level or model-based estimates. However, they do include estimates for the mid-Atlantic division and the Northeast region. Estimates based on 2006–2010 data may differ from previously published estimates. Tables and estimates based only on 2011 data are unaffected by these data errors. All affected tables, that is, tables with estimates based on 2006–2010 data, contain a note to indicate this to the user.

Caution is advised when comparing data from older reports with data from more recent reports that are based on corrected data files. As discussed above, comparisons of estimates for Pennsylvania, Maryland, the mid-Atlantic division, and the Northeast region are of most concern, whereas comparisons of national data or data for other States and regions are essentially still valid. A selected set of corrected versions of reports and tables have been produced.
In particular, a set of modified detailed tables that include revised 2006–2010 estimates for the

mid-Atlantic division and the Northeast region for certain key measures have been released. Given the change noted above, comparisons between unrevised 2006–2010 estimates and estimates based on 2011 or 2012 data for the areas of most concern are not recommended.

4. Missingness

4.1 Potential Estimation Bias Due to Missingness

In the 2012 National Survey on Drug Use and Health (NSDUH), many variables, including core drug and demographic variables, had missing item response values imputed. See the 2012 NSDUH imputation report (Frechtel et al., 2014) for further details. However, the missing item responses of many other variables were not imputed, and these missing responses may lead to biased estimates in the 2012 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b). In addition, another source of potential uncertainty about some estimates may occur because of the way unknown item responses (e.g., blank, "don't know," "refused") were actually coded for different variables. For example, some recoded variables (i.e., variables created from one or more source variables) classified unknown item responses in the source variable(s) as missing values, whereas others did not. See Ruppenkamp, Emrich, Aldworth, Hirsch, and Foster (2006) for further details.

Recall from Section 3 that prevalence rates are defined as the proportions of the population who exhibit characteristics of interest. Let $\hat{p}_d$ represent the estimated prevalence rate of interest for domain *d*, with $\hat{p}_d$ defined as

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

where $\hat{Y}_d$ = estimated number of persons exhibiting the characteristic of interest in domain *d*, and $\hat{N}_d$ = estimated population total for domain *d*.

The variable defining the characteristic of interest (e.g., illicit drug use) is referred to as the *analysis* variable, and the variable defining the domain of interest (e.g., receipt of past year mental health treatment/counseling) is referred to as the *domain* variable. Suppose that the analysis variable has all its missing values imputed, but the domain variable does not employ the imputation of missing values. In such cases, the estimates $\hat{N}_d$ and $\hat{Y}_d$ may be negatively biased, and the $\hat{p}_d$ estimates also may be biased. To see this, suppose that the domain variable has *D* levels, and define

$$\hat{N} = \sum_{d=1}^{D} \hat{N}_d + \hat{N}_m,$$

where $\hat{N}$ = estimated population total, $\hat{N}_d$ = estimated population total for domain *d*, $d = 1, 2, ..., D$, and $\hat{N}_m$ = estimated population total corresponding to the missing values of the domain variable. Thus, if $\hat{N}_m$ is positive (i.e., there are missing domain-variable responses), then

at least one of the $\hat{N}_d$ estimates will be negatively biased. The presence of negative bias in at least one of the $\hat{Y}_d$ estimates can be similarly demonstrated if $\hat{Y}_m$ is positive, where $\hat{Y}_m$ = the estimated number of persons exhibiting the characteristic of interest and corresponding to the missing values of the domain variable. If either of $\hat{N}_m$ and $\hat{Y}_m$ is positive, then $\hat{p}_d$ may be biased by some unknown amount.

In the 2012 detailed tables (CBHSQ, 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b), potential bias in the $\hat{N}_d$, $\hat{Y}_d$, or $\hat{p}_d$ estimates was not treated, although footnotes included on the tables provide detailed information about which estimates were based on or excluded missing values. This problem may be illustrated by the following example, which corresponds to information presented in Tables 2.9A and 2.9B of the 2012 mental health detailed tables (CBHSQ, 2013b).

Mental health Table 2.9A presents estimates of the past year use of several types of illicit drugs among persons aged 12 to 17 for 2011 and 2012. These analysis variables are grouped into a two-level domain variable that is categorized according to whether a respondent had a past year major depressive episode (MDE). In 2012, mental health Table 3.2A shows the population estimate of persons aged 12 to 17 as approximately 24,933,000. However, the subdomain population estimates summed to approximately 24,400,000, resulting in an estimate of $\hat{N}_m$ = 533,000 (approximately 2.1 percent of the total population). This number represents the estimated population not assigned to either domain. This negative bias can extend to various analysis variables, such as "Illicit Drugs." In 2012, the total estimate of persons aged 12 to 17 who used illicit drugs in the past year was approximately 4,447,000. However, the estimates of persons aged 12 to 17 who used illicit drugs in the past year among the valid subdomains (where past year MDE status was not missing) summed to 4,370,000, resulting in an estimate of $\hat{Y}_m$ = 98,000 (approximately 2.2 percent of the total population).

Mental health Table 2.9B presents prevalence estimates of the past year use of several types of illicit drugs among persons aged 12 to 17 for 2011 and 2012. Because $\hat{N}_m$ is positive and $\hat{Y}_m$ is positive for the analysis variable, "Illicit Drugs," the prevalence estimates for this variable may be biased by some unknown amount across the two domains. The 2012 prevalence estimates reported in mental health Table 2.9B for youths who had or did not have past year MDE are 34.0 and 16.3 percent, respectively. It can be shown that the approximate range of possible bias values for each of these estimates is as follows: between -5.60 and 2.79 percent and between -0.31 and 0.37 percent, respectively.

4.2 Variance Estimation in the Presence of Missingness

SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2008) uses the number of strata and number of primary sampling units (PSUs) in its variance calculations, even if there are some PSUs in which a variable is entirely missing for all sample members associated with that PSU (RTI International, 2008). The rationale behind this approach

is that there may be individuals in the target population who have nonmissing values in PSUs where no sample members have nonmissing values.

To illustrate how this is operationalized in SUDAAN, consider the following example. Suppose there is interest in calculating the mean of some variable (say, X), but there are missing values associated with X. SUDAAN then creates an internal subpopulation indicator variable (say, δ), where $\delta = 1$ if X is not missing, and $\delta = 0$ if X is missing. Then SUDAAN internally calculates the mean and variance of X by using δX.

For the variance estimator based on the Taylor series linearization approach, one of the terms in the variance estimator consists of the sum of squared deviations of PSU-level totals about their stratum-level means, divided by the number of PSUs in the strata minus 1. Therefore, if SUDAAN encounters an incorrect number of PSUs within a stratum, then this term is incorrectly calculated. In addition, if there is only one PSU in a stratum, then the denominator for the variance term associated with that stratum becomes zero, and this causes the overall variance estimate to return an error message in SUDAAN. Hence, PSUs (associated with missing values) should never be excluded from an input file.

5. Sampling Error

As were the prevalence rates, all of the variance estimates for prevalence (including those for prevalence based on annual averages from combined data) were calculated using a method in SUDAAN® that is unbiased for linear statistics. This method is based on multistage clustered sample designs where the first-stage (primary) sampling units are drawn with replacement.

Because of the complex nature of the sampling design for the National Survey on Drug Use and Health (NSDUH) (specifically the use of stratified-clustering sampling), key nesting variables were created for use in SUDAAN to capture explicit stratification and to identify clustering. Starting with the 2005 NSDUH, a change was made in the way the key nesting variables were defined. Each State sampling region (SSR) appears in a different variance estimation stratum every quarter. This method had the effect of assigning the regions to strata in a pseudo-random fashion while ensuring that each stratum consists of four SSRs from four different States.

Two replicates per year were defined within each variance stratum (VEREP). Each variance replicate consists of four segments, one for each quarter of data collection. One replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The other replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. A segment stays in the same VEREP for the 2 years it is in the sample. This simplifies computing standard errors (SEs) for estimates based on combined data from adjacent survey years.

Although the SEs of estimates of means and proportions can be calculated appropriately in SUDAAN using a Taylor series linearization approach, SEs of estimates of totals may be underestimated in situations where the domain size is poststratified to data from the U.S. Census Bureau. Because of this underestimation, alternatives for estimating SEs of totals were implemented in all of the 2012 detailed tables (CBHSQ, 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b), where appropriate.

Estimates of means or proportions, $\hat{p}_d$, such as drug use prevalence rates for a domain *d*, can be expressed as a ratio estimate:

$$\hat{p}_d = \frac{\hat{Y}_d}{\hat{N}_d},$$

where $\hat{Y}_d$ is a linear statistic estimating the number of substance users in the domain *d*, and $\hat{N}_d$ is a linear statistic estimating the total number of persons in domain *d* (both users and nonusers). The SUDAAN software package is used to calculate direct estimates of $\hat{Y}_d$ and $\hat{N}_d$ and also can be used to estimate their respective SEs. A Taylor series approximation method implemented in SUDAAN provides estimates for $\hat{p}_d$ and its SE.

When the domain size, $\hat{N}_d$, is free of sampling error, an appropriate estimate of the SE for the total number of substance users is

$$\mathrm{SE}(\hat{Y}_d) = \hat{N}_d \mathrm{SE}(\hat{p}_d).$$

This approach is theoretically correct when the domain size estimates, $\hat{N}_d$, are among those forced to match their respective U.S. Census Bureau population estimates through the weight calibration process.[11] In these cases, $\hat{N}_d$ is not subject to a sampling error induced by the NSDUH design.

For estimated domain totals, $\hat{Y}_d$, where $\hat{N}_d$ is not fixed (i.e., where domain size estimates are not forced to match the U.S. Census Bureau population estimates), this formulation still may provide a good approximation if it can be assumed that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$. This is a reasonable assumption for most cases in this study.

For various subsets of estimates, the above approach yielded an underestimate of the variance of a total because $\hat{N}_d$ was subject to considerable variation. In 2000, an approach was implemented to reflect more accurately the effects of the weighting process on the variance of total estimates. This approach consisted of calculating SEs of totals for all estimates in a particular detailed table using the formula above when a majority of estimates in a table were among domains in which $\hat{N}_d$ was fixed during weighting or if it could be assumed that the sampling variation in $\hat{N}_d$ was negligible. Detailed tables in which the majority of estimates were among domains where $\hat{N}_d$ was subject to considerable variability were calculated directly in SUDAAN.

To improve on the accuracy of the SEs, a "mixed" method approach was implemented. This method was applied to selected tables in the 2004 NSDUH, and it was implemented across all tables starting with the 2005 NSDUH and continuing in the 2012 NSDUH. This approach assigns the method of SE calculation to domains within tables so that all estimates among a select set of domains with fixed $\hat{N}_d$ were calculated using the formula above, and all other estimates were calculated directly in SUDAAN, regardless of other estimates within the same table. The set of domains considered controlled (i.e., those with a fixed $\hat{N}_d$) was restricted to main effects and two-way interactions to maintain continuity between years. Domains consisting of three-way interactions may be controlled in 1 year but not necessarily in preceding or subsequent years. The use of such SEs did not affect the SE estimates for the corresponding proportions presented in the same sets of tables because all SEs for means and proportions are calculated directly in SUDAAN. Appendix A contains example SAS and SUDAAN code that

[11] For more information on the sampling weight calibration in the 2012 NSDUH, see Chen et al. (2014).

demonstrates how to compute SEs of proportions as well as both types of SEs of totals (controlled or uncontrolled).

Table 3 contains a list of domains with a fixed $\hat{N}_d$ for the restricted use data file.[12] This table includes both the main effects and two-way interactions and may be used to identify the method of SE calculation employed for estimates of totals in the 2012 detailed tables (CBHSQ, 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b). For example, Table 1.23 of the 2012 detailed tables presents estimates of illicit drug use among persons aged 18 or older within the domains of gender, Hispanic or Latino (referred to as "Hispanic" hereafter) origin and race, education, and current employment. Estimates among the total population (age main effect), males and females (age by gender interaction), and Hispanics and non-Hispanics (age by Hispanic origin interaction) were treated as controlled in this table, and the formula above was used to calculate the SEs. The SEs for all other estimates, including white and black or African American (age by Hispanic origin by race interaction), were calculated directly from SUDAAN. It is important to note that estimates presented in the 2012 detailed tables and 2012 mental health detailed tables for racial groups are among non-Hispanics, unless noted otherwise. For instance, the domain for whites is actually non-Hispanic whites and is therefore a two-way interaction. Although not reported on in the 2012 detailed tables or the 2012 mental health detailed tables, additional geographic interactions are also treated as domains with fixed $\hat{N}_d$ for other NSDUH analyses. Similar to geographic region and division, a State is considered a controlled domain, and two-way interactions with State and gender, Hispanic origin, quarter, and age group (12-17, 18-25, and 26 or older) are all treated as domains with fixed $\hat{N}_d$.

[12] See the estimation of totals section in the public use file introduction for a list of domains with fixed $\hat{N}_d$ for the 2012 public use file (CBHSQ/SAMHSA, 2013).

Table 3. Demographic and Geographic Domains Forced to Match Their Respective U.S. Census Bureau Population Estimates through the Weight Calibration Process, 2012

Main Effects	Two-Way Interactions
Age Group	
12-17	
18-25	**Age Group × Gender**
26-34	(e.g., Males Aged 12 to 17)
35-49	
50-64	
65 or Older	**Age Group × Hispanic Origin**
All Combinations of Groups Listed Above[1]	(e.g., Hispanics or Latinos Aged 18 to 25)
Gender	
Male	
Female	**Age Group × Race**
Hispanic Origin	(e.g., Whites Aged 26 or Older)
Hispanic or Latino	
Not Hispanic or Latino	
Race	**Age Group × Geographic Region**
White	(e.g., Persons Aged 12 to 25 in the Northeast)
Black or African American	
Geographic Region	
Northeast	**Age Group × Geographic Division**
Midwest	(e.g., Persons Aged 65 or Older in New England)
South	
West	
Geographic Division	**Gender × Hispanic Origin**
New England	(e.g., Not Hispanic or Latino Males)
Middle Atlantic	
East North Central	
West North Central	**Hispanic Origin × Race**
South Atlantic	(e.g., Not Hispanic or Latino Whites)
East South Central	
West South Central	
Mountain	
Pacific	

[1] Combinations of the age groups (including but not limited to 12 or older, 18 or older, 26 or older, 35 or older, and 50 or older) also were forced to match their respective U.S. Census Bureau population estimates through the weight calibration process.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

6. Degrees of Freedom

To determine whether the observed difference between estimates is statistically significant, the degrees of freedom (*df*) are needed to locate the corresponding probability level (*p* value) of the test statistic. The test statistic is computed from the sample data and represents a numerical summary of the difference between the estimates under consideration; it is a random variable that has a predetermined distribution (such as Student's *t*, chi-square, or *F*). The degrees of freedom characterize the amount of variation expected in the estimation of sampling error and are used in conjunction with the test statistic to determine probabilities and evaluate statistical significance. In statistics, the number of degrees of freedom refers to the number of independent units of information in a sample relevant to the estimation of a parameter or calculation of a statistic. In general, the degrees of freedom of a parameter estimate is equal to the number of independent observations that go into the estimate minus the number of other parameters that need to be estimated as an intermediate step. The degrees of freedom are also used to compute the confidence intervals (CIs) discussed in Section 8. The upper and lower limits of the CIs are defined by a constant value that is chosen to yield a level of confidence based on the degrees of freedom.

Starting in 2005, there was a change in definition to the variance estimation strata for the National Survey on Drug Use and Health (NSDUH). This change in definition, which was applied to the 2005–2012 NSDUHs, has the effect of increasing the number of degrees of freedom for State-level estimates while preserving the number of degrees of freedom for national estimates (900). The degrees of freedom are calculated as the number of primary sampling units (variance replicates) minus the number of strata for the data being analyzed. Because the NSDUH sample design provides for estimates by State in all 50 States plus the District of Columbia, States may be viewed as the first level of stratification. When producing NSDUH estimates on the national level, including estimates based on annual averages from combined data, there are 900 degrees of freedom. If an analysis only involves certain States, the degrees of freedom change depending on whether the State is a large sample or small sample State. The large sample States (i.e., California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) have 192 degrees of freedom because each large State is in 192 strata. All of the other States (i.e., the small sample States, which include the District of Columbia) have 48 degrees of freedom because each small State is in 48 different strata. Note that the 2012 detailed tables (CBHSQ, 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b) use 900 degrees of freedom for all estimates, including those for geographic regions and divisions.

When testing between groups that have differing degrees of freedom, the minimum degrees of freedom from the groups are used, which makes for a more conservative test. For all nonstandard geographic areas (e.g., geographic areas other than region and division) or when multiple years are grouped for analyses, the degrees of freedom are calculated outside of SUDAAN, and this value is entered manually into SUDAAN for use in testing. Appendix A contains an example that demonstrates how to define the degrees of freedom within the SUDAAN (RTI International, 2008) procedure to compute design-based estimates.

For an analysis of a group of States, the degrees of freedom would be less than or equal to the sum of the degrees of freedom for each individual State due to overlap of strata. The

specific number of degrees of freedom can be computed by counting the unique values of VESTR (variance estimation [pseudo] stratum) for the particular geographic area of interest. For these types of specific State analyses (or other subpopulations of interest), the degrees of freedom can be specifically indicated in SUDAAN (RTI International, 2008); otherwise, the degrees of freedom are computed using the entire dataset. Similar methods can be used to compute appropriate degrees of freedom for any geographic region comprising counties or States as well. As noted previously, the detailed tables and mental health detailed tables deviate from this rule and use 900 degrees of freedom for all estimates (including census region and division estimates and estimates for all years including pooled years). The technique of counting the number of unique values of VESTR can also be used for analyses combining survey data across years.

7. Statistical Significance of Differences

Once the degrees of freedom have been determined, various methods used to compare prevalence estimates may be employed. This section describes some of these methods. Customarily, the observed difference between estimates is evaluated in terms of its statistical significance. Statistical significance is based on the *p* value of the test statistic and refers to the probability that a difference as large as that observed would occur due to random variability in the estimates if there were no differences in the prevalence rates being compared. The significance of observed differences is generally reported at the .05 and .01 levels when the *p* value is defined as less than or equal to the designated significance level.

Significance tests were conducted on differences between prevalence estimates from the 2012 National Survey on Drug Use and Health (NSDUH) and previous years of NSDUH back to 2002. Due to survey design changes implemented in 2002, data from the 2002–2012 NSDUHs should not be compared with data from earlier survey years. Significance tests also were conducted on differences of prevalence estimates between combined 2009–2010 survey data and combined 2011–2012 survey data. Within-year tests were conducted on differences between prevalence estimates for various populations (or subgroups) of interest using data from the 2012 survey.

When comparing prevalence estimates, one can test the null hypothesis (no difference between rates) against the alternative hypothesis (there is a difference in prevalence rates) using the standard *t* test (with the appropriate degrees of freedom) for the difference in proportions test, expressed as

$$t_{df} = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\mathrm{var}(\hat{p}_1) + \mathrm{var}(\hat{p}_2) - 2\,\mathrm{cov}(\hat{p}_1, \hat{p}_2)}},$$

where *df* = the appropriate degrees of freedom, $\hat{p}_1$ = first prevalence estimate, $\hat{p}_2$ = second prevalence estimate, var($\hat{p}_1$) = variance of first prevalence estimate, var($\hat{p}_2$) = variance of second prevalence estimate, and cov($\hat{p}_1, \hat{p}_2$) = covariance between $\hat{p}_1$ and $\hat{p}_2$. Note that the first and second prevalence estimates may take the form of prevalence estimates from two different survey years (e.g., 2011 and 2012, respectively), prevalence estimates from sets of combined survey data (e.g., 2009–2010 annual averages and 2011–2012 annual averages, respectively), or prevalence estimates for populations of interest within a single survey year.

Under the null hypothesis, the test statistic *t* is a random variable that asymptotically follows a *t*-distribution. Therefore, calculated values of *t*, along with the appropriate degrees of freedom, can be used to determine the corresponding probability level (i.e., *p* value). Whether testing for differences between years or from different populations within the same year, the covariance term in the formula for *t* will, in general, not be equal to zero. SUDAAN is used to compute estimates of *t* along with the associated *p* values such that the covariance term is calculated by taking the sample design into account (RTI International, 2008). A similar procedure and formula for *t* are used for estimated totals; however, it should be noted that

because it was necessary to calculate the standard error (SE) outside SUDAAN for domains forced by the weighting process to match their respective U.S. Census Bureau population estimates, the corresponding test statistics also were computed outside SUDAAN. SAS and SUDAAN examples showing the computational methods for generating *p* values of estimates of *t* and estimated totals can be found in Appendix A.

Under the null hypothesis, the test statistic with known variances asymptotically follows a standard normal (*Z*) distribution. However, because the variances of the test statistic are estimated, its distribution is more accurately described by the *t*-distribution for finite sample sizes. A sufficiently large sample size is required for the asymptotic properties to take effect, and this is usually determined through the suppression criteria applied to the estimates (see Section 10). As the degrees of freedom approach infinity, the *t*-distribution approaches the *Z* distribution. That is, because most of the statistical tests performed have 900 degrees of freedom, the *t* tests performed produce approximately the same numerical results as if a *Z* test had been performed.

When comparing population subgroups defined by three or more levels of a categorical variable, log-linear chi-square tests of independence of the subgroup and the prevalence variables were conducted first to control the error level for multiple comparisons. If Shah's Wald *F* test (transformed from the standard Wald chi-square) indicated overall significant differences, the significance of each particular pairwise comparison of interest was tested using SUDAAN analytic procedures to properly account for the sample design (RTI International, 2008).

If SUDAAN is not available to compute the significance testing, using published estimates can provide similar testing results. When comparing prevalence rates shown with SEs, independent *t* tests for the difference of proportions can be performed and usually will provide the same results as tests performed in SUDAAN. However, where the *p* value is close to the predetermined level of significance, results may differ for two reasons: (1) the covariance term is included in the SUDAAN tests, whereas it is not included in independent *t* tests; and (2) the reduced number of significant digits shown in the published estimates may cause rounding errors in the independent *t* tests. Although not generated in all NSDUH publications, some publications do include sampling error in the form of 95 percent confidence intervals (CIs). In terms of testing for differences between prevalence rates shown with 95 percent CIs, it is important to note that two overlapping 95 percent CIs do not imply that their rates are statistically equivalent at the 5 percent level of significance. For additional information, see Schenker and Gentleman (2001) and Payton, Greenstone, and Schenker (2003).

8. Confidence Intervals

In some National Survey on Drug Use and Health (NSDUH) publications, sampling error has been quantified using 95 percent confidence intervals (CIs). Frequently, NSDUH estimates are small percentages (i.e., are close to 0), and in that case, a logit transformation of the estimate provides favorable properties. For example, the logit transformation yields asymmetric interval boundaries between 0 and 1 that are more balanced with respect to the true probability that the true value falls below or above the interval boundaries. This is partly because for values close to 0, the distribution of a logit-transformed estimate approximates the normal distribution more closely than the standard estimate. Standard symmetric CIs for small proportions may also lead to the undesirable result of a lower CI limit that is less than 0.

To illustrate the method, let the proportion P_d represent the true prevalence rate for a particular analysis domain *d*. Then the logit transformation of P_d , commonly referred to as the "log odds," is defined as

$$L = \ln[P_d / (1 - P_d)],$$

where "1n" denotes the natural logarithm.

Letting $\hat{p}_d$ be the estimate of the domain proportion, the log odds estimate becomes

$$\hat{L} = \ln[\hat{p}_d /(1 - \hat{p}_d)].$$

The lower and upper confidence limits of *L* are formed as

$$A = \hat{L} - K\left[\frac{\sqrt{\mathrm{var}(\hat{p}_d)}}{\hat{p}_d(1-\hat{p}_d)}\right],$$
$$B = \hat{L} + K\left[\frac{\sqrt{\mathrm{var}(\hat{p}_d)}}{\hat{p}_d(1-\hat{p}_d)}\right],$$

where $\mathrm{var}(\hat{p}_d)$ is the variance estimate of $\hat{p}_d$, the quantity in brackets is a first-order Taylor series approximation of the standard error of $\hat{L}$, and *K* is the critical value of the *t*-distribution associated with a specified level of confidence and degrees of freedom (*df*). For example, to produce 95 percent confidence limits for national estimates, the value of *K* would be 1.96 based on 900 degrees of freedom (similarly, for large States, *K* would be 1.97 based on 192 degrees of freedom, and for small States, *K* would be 1.98 based on 48 degrees of freedom).

Although the distribution of the logit-transformed estimate, $\hat{L}$, is asymptotically normal, the variance term in the CI is estimated, and a critical value from the *t*-distribution is therefore appropriate when calculating CIs. A sufficiently large sample size is required for the asymptotic

properties to take effect, and this is usually determined through the suppression criteria applied to the estimates (see Section 10).

Applying the inverse logit transformation to A and B above yields a CI for $\hat{p}_d$ as follows:

$$\hat{p}_{d,lower} = \frac{1}{1+\exp(-A)},$$

$$\hat{p}_{d,upper} = \frac{1}{1+\exp(-B)},$$

where "exp" denotes the inverse log transformation. The lower and upper CI endpoints for percentage estimates are obtained by multiplying the lower and upper endpoints of $\hat{p}_d$ by 100.

The CI for the estimated domain total, $\hat{Y}_d$, as estimated by

$$\hat{Y}_d = \hat{N}_d \cdot \hat{p}_d,$$

is obtained by multiplying the lower and upper limits of the proportion CI by $\hat{N}_d$. For domain totals $\hat{Y}_d$, where $\hat{N}_d$ is not fixed, the CI approximation assumes that the sampling variation in $\hat{N}_d$ is negligible relative to the sampling variation in $\hat{p}_d$.

9. Incidence Estimates

In epidemiological studies, incidence is defined as the number of new cases of a disease occurring within a specific period of time. Similarly, in substance use studies, incidence refers to the first use of a particular substance.

Starting with the 2004 National Survey on Drug Use and Health (NSDUH) data, the evaluation of trends in the initiation of drug use was presented by estimates of past year drug use incidence or initiation (i.e., the number of users whose first use was within the 12 months before their interview date). This incidence measure, termed "past year initiation," is determined by self-reported past year use, age at first use, year and month of most recent new use, and the interview date.

Since 1999, the NSDUH questionnaire allowed for the collection of year and month of first use for recent initiates (i.e., persons who used a particular substance for the first time in a given survey year). Month, day, and year of birth also were obtained directly or imputed for item nonrespondents as part of the data processing. In addition, the questionnaire call record provided the date of the interview. By imputing a day of first use within the year and month of first use, a specific date of first use, $t_{fu,d,i}$, can be used for estimation purposes.

Past year initiation among persons using a substance in the past year can be viewed as an indicator variable defined as follows:

$$I_{(Past\ Year\ Initiate)}(i) = \begin{cases} 1 & \text{if } \left(DOI_i MOI_i YOI_i - t_{fu,d,i}\right) \leq 365 \\ 0 & \text{otherwise} \end{cases},$$

where DOI_i, MOI_i, and YOI_i denote the day, month, and year of the interview for person *i*, respectively, and $t_{fu,d,i}$ denotes the date of first use associated with person *i*.

The calculation of past year initiation does not take into account whether the respondent initiated substance use while a resident of the United States. This method of calculation has little effect on past year estimates and provides direct comparability with other standard measures of substance use because the populations of interest for the measures will be the same (i.e., both measures examine all possible respondents and do not restrict to those only initiating substance use in the United States).

One important note for incidence estimates is the relationship between a main substance category and subcategories of substances (e.g., illicit drugs would be a main category and inhalants and marijuana would be examples of subcategories in relation to illicit drugs). For most measures of substance use, any member of a subcategory is by necessity a member of the main category (e.g., if a respondent is a past month user of a particular drug, then he or she is also a past month user of illicit drugs in general). However, this is not the case with regard to incidence statistics. Because an individual can only be an initiate of a particular substance category (main or sub) a single time, a respondent with lifetime use of multiple substances may not,

by necessity, be included as an initiate of a main category, even if he or she were an initiate for a particular subcategory because his or her first initiation of other substances could have occurred earlier.

In addition to estimates of the number of persons initiating use of a substance in the past year, estimates of the mean age of past year first-time users of these substances were computed. Unless specified otherwise, estimates of the mean age at initiation in the past 12 months have been restricted to persons aged 12 to 49 so that the mean age estimates reported are not influenced by those few respondents who were past year initiates at age 50 or older. As a measure of central tendency, means are influenced heavily by the presence of extreme values in the data, and this constraint should increase the utility of these results to health researchers and analysts by providing a better picture of the substance use initiation behaviors among the civilian, noninstitutionalized population in the United States. This constraint was applied only to estimates of mean age at first use and does not affect estimates of incidence.

Because NSDUH is a survey of persons aged 12 or older at the time of the interview, younger individuals in the sample dwelling units are not eligible for selection into the NSDUH sample. Some of these younger persons may have initiated substance use during the past year. As a result, past year initiate estimates suffer from undercoverage when one can think of the estimates as reflecting all initial users regardless of current age. For earlier years, data can be obtained retrospectively based on the age at and date of first use. As an example, persons who were 12 years old on the date of their interview in the 2012 survey may have reported initiating use of cigarettes between 1 and 2 years ago; these persons would have been past year initiates reported in the 2011 survey had persons who were 11 years old on the date of the 2011 interview been allowed to participate in the survey. Similarly, estimates of past year use by younger persons (aged 10 or younger) can be derived from the current survey, but they apply to initiation in prior years—not the survey year.

To get an impression of the potential undercoverage in the current year, reports of substance use initiation reported in 2012 by persons aged 12 or older were estimated for the years in which these persons would have been 1 to 11 years younger. These estimates do not necessarily reflect behavior by persons who were 1 to 11 years younger in 2012. Instead, the data for the 11-year-olds reflect initiation in the year before the 2012 survey, the data for the 10-year-olds reflect behavior between the 12th and 23rd month before the 2012 survey, and so on. A rough way to adjust for the difference in the years that the estimate pertains to without considering changes to the population is to apply an adjustment factor to each age-based estimate of past year initiates. The adjustment factor can be based on a ratio of lifetime users aged 12 to 17 in 2012 to the same estimates for the prior applicable survey year. To illustrate the calculation, consider past year use of alcohol. In the 2012 survey, 72,063 persons who were 12 years old in 2011 were estimated to have initiated use of alcohol between 1 and 2 years earlier. These persons would have been past year initiates in the 2011 survey conducted on the same dates had the 2011 survey covered younger persons. The estimated number of lifetime users currently aged 12 to 17 was 8,067,487 for 2012 and 8,610,370 for 2011, indicating fewer overall initiates of alcohol use among persons aged 17 or younger in 2012. Thus, an adjusted estimate of initiation of alcohol use by persons who were 11 years old in 2012 is given by

$$(Estimated\ Past\ Year\ Initiates\ Aged\ 11)_{2011} \times \frac{(Estimated\ Lifetime\ Users\ Aged\ 12\ to\ 17)_{2012}}{(Estimated\ Lifetime\ Users\ Aged\ 12\ to\ 17)_{2011}}\ .$$

Numerically, this yielded an adjusted estimate of 67,519 persons who were 11 years old on a 2012 survey date and initiated use of alcohol in the past year:

$$72,063 \times \frac{8,067,487}{8,610,370} = 67,519$$

A similar procedure was used to adjust the estimated number of past year initiates among persons who would have been 10 years old on the date of the interview in 2010 and for younger persons in earlier years. The overall adjusted estimate for past year initiates of alcohol use by persons aged 11 or younger on the date of the interview was 150,620, or about 3.3 percent of the estimate based on past year initiation by persons aged 12 or older only (150,620 ÷ 4,589,405 = 0.0328).

Based on similar analyses, the estimated undercoverage of past year initiates aged 11 or younger was about 3.2 percent for cigarettes, about 1.1 percent for marijuana, and about 18.8 percent for inhalants.

The undercoverage of past year initiates aged 11 or younger also affects the mean age-at-first-use estimate. An adjusted estimate of the mean age at first use was calculated using a weighted estimate of the mean age at first use based on the current survey and the numbers of persons aged 11 or younger in the past year obtained in the aforementioned analysis for estimating undercoverage of past year initiates. Analysis results showed that the mean age at first use was changed from 17.4 to 17.1 for alcohol, from 17.8 to 17.5 for cigarettes, from 17.9 to 17.8 for marijuana, and from 16.9 to 15.4 for inhalants. The decreases reported above are comparable with results generated in prior survey years.

10. Suppression of Estimates with Low Precision

Direct survey estimates that were considered to be unreliable because of unacceptably large sampling errors were not reported, but rather were noted by an asterisk (*). The criteria used to assess the need to suppress direct survey estimates were based on prevalence (for proportion estimates), the relative standard error (RSE) (defined as the ratio of the standard error [SE] over the estimate), nominal (actual) sample size, and effective sample size for each estimate.

Proportion estimates ($\hat{p}$), or rates, within the range $0 < \hat{p} < 1$, and corresponding estimated numbers of users were suppressed if

$$\mathrm{RSE}[-\ln(\hat{p})] > .175 \text{ when } \hat{p} \le .5$$

or

$$\mathrm{RSE}[-\ln(1-\hat{p})] > .175 \text{ when } \hat{p} > .5 .$$

Based on a first-order Taylor series approximation of RSE $[-\ln(\hat{p})]$ and RSE $[-\ln(1-\hat{p})]$, the following equation was derived and used for computational purposes when applying a suppression rule dependent on effective sample sizes:

$$\frac{\mathrm{SE}(\hat{p})/\hat{p}}{-\ln(\hat{p})} > .175 \text{ when } \hat{p} \le .5,$$

or

$$\frac{\mathrm{SE}(\hat{p})/(1-\hat{p})}{-\ln(1-\hat{p})} > .175 \text{ when } \hat{p} > .5 .$$

The separate formulas for $\hat{p} \le .5$ and $\hat{p} > .5$ produce a symmetric suppression rule; that is, if $\hat{p}$ is suppressed, $1-\hat{p}$ will be suppressed as well. See Figure 1 for a graphical representation of the required minimum effective sample sizes as a function of the proportion estimated. When $.05 < \hat{p} < .95$, the symmetric properties of the rule produce local minimum effective sample sizes at $\hat{p} = .2$ and again at $\hat{p} = .8$, such that an effective sample size of greater than 50 is required; this means that estimates would be suppressed for these values of $\hat{p}$ unless the effective sample sizes were greater than 50. Within this same interval of $.05 < \hat{p} < .95$, a local maximum effective sample size of 68 is required at $\hat{p} = .5$. So, to simplify requirements and maintain a conservative suppression rule, estimates of $\hat{p}$ between .05 and .95, which had effective sample sizes below 68, were suppressed.

The effective sample size for a domain is a function of the nominal sample size and the design effect (i.e., nominal sample size/design effect). During the original development of this suppression rule, the design effect was calculated outside SUDAAN (RTI International, 2008) in SAS. Since the 2005 National Survey on Drug Use and Health (NSDUH) analysis, the direct

SUDAAN design effect was used to provide a more precise and accurate reflection of the design effect (due to the removal of several possible rounding errors) when compared with the SAS method used in the past. The differences between the direct SUDAAN design effects and the SAS-calculated design effects only occur at approximately the tenth decimal place or later; however, previously published estimates that were on the borderline of being suppressed or unsuppressed due to the effective sample size suppression rule may potentially change from suppressed to unsuppressed, or vice versa.

Figure 1. Required Effective Sample in the 2012 NSDUH as a Function of the Proportion Estimated

In addition, a minimum nominal sample size suppression criterion (n = 100) that protects against unreliable estimates caused by small design effects and small nominal sample sizes was employed. Table 4 shows a formula for calculating design effects. Prevalence estimates also were suppressed if they were close to 0 or 100 percent (i.e., if $\hat{p} < .00005$ or if $\hat{p} \geq .99995$).

Beginning with the 1991 survey, the suppression rule for proportions based on RSE $[-\ln(\hat{p})]$ described above replaced an older rule in which data were suppressed whenever RSE ($\hat{p}$) > .5. This rule was changed because the older rule imposed a very stringent application for small $\hat{p}$, but a very lax application for large $\hat{p}$. The new rule ensured a more uniformly stringent application across the whole range of $\hat{p}$ (i.e., from 0 to 1). The old rule also was asymmetric in the sense that suppression only occurred in terms of $\hat{p}$; that is, there was no complementary rule for (1 – $\hat{p}$), which the new suppression rules now account for.

Estimates of totals were suppressed if the corresponding prevalence rates were suppressed. Estimates of means not bounded between 0 and 1 (e.g., mean age at first use, mean number of drinks consumed) were suppressed if the RSEs of the estimates were larger than .5 or if the sample sizes were smaller than 10 respondents. In these cases, the totals (e.g., total number of drinks consumed) were suppressed if the corresponding mean estimates were suppressed.

Section 4 demonstrates an exception to the rule that the totals are suppressed when their corresponding means are suppressed. Some tables in Section 4 show estimates of incidence among different populations. Specifically, these Section 4 tables display the number of initiates among three different populations: the total population, persons at risk for initiation, and past year users. In these tables, some mean estimates may be suppressed while the total estimate is not suppressed. When at least one mean estimate in the table is not suppressed, one can assume that the numerator (or total estimate) is not the cause for the suppression and the total estimate will not be suppressed. In contrast, when all mean estimates are suppressed, the total will also be suppressed.

Tables that show sample sizes and population counts do not incorporate the suppression rule for several reasons. One reason is that no mean is associated with these estimates; thus, most of the components of the suppression criteria are not applicable. Also, because no behavior associated with the numbers is displayed, there is no risk of behavior disclosure.

The suppression criteria for various NSDUH estimates are summarized in Table 4, and sample SAS code demonstrating how to implement these rules can be found in Appendix A.

Table 4. Summary of 2012 NSDUH Suppression Rules

Estimate	Suppress if:
Prevalence Rate, $\hat{p}$, with Nominal Sample Size, *n*, and Design Effect, *deff* $\left(deff = \frac{n[\mathrm{SE}(\hat{p})]^2}{\hat{p}(1-\hat{p})} \right)$	(1) The estimated prevalence rate, $\hat{p}$, is < 0.00005 or ≥ 0.99995, or (2) $\frac{\mathrm{SE}(\hat{p}) \ / \ \hat{p}}{-\ln(\hat{p})} > 0.175$ when $\hat{p} \le 0.5$, or $\frac{\mathrm{SE}(\hat{p}) \ / \ (1 - \hat{p})}{-\ln(1 - \hat{p})} > 0.175$ when $\hat{p} > 0.5$, or (3) Effective $n < 68$, where Effective $n = \frac{n}{deff} = \frac{\hat{p}(1-\hat{p})}{\left[\mathrm{SE}(\hat{p})\right]^2}$, or (4) $n < 100$. Note: The rounding portion of this suppression rule for prevalence rates will produce some estimates that round at one decimal place to 0.0 or 100.0 percent but are not suppressed from the tables.
Estimated Number (Numerator of $\hat{p}$)	The estimated prevalence rate, $\hat{p}$, is suppressed. Note: In some instances when $\hat{p}$ is not suppressed, the estimated number may appear as a 0 in the tables. This means that the estimate is greater than 0 but less than 500 (estimated numbers are shown in thousands). Note: In some instances when totals corresponding to several different means that are displayed in the same table and some, but not all, of those means are suppressed, the totals will not be suppressed. When all means are suppressed, the totals will also be suppressed.
Means not bounded between 0 and 1 (i.e., Mean Age at First Use, Mean Number of Drinks), $\bar{x}$, with Nominal Sample Size, *n*	(1) $\mathrm{RSE}(\bar{x}) > 0.5$, or (2) $n < 10$.

deff = design effect; RSE = relative standard error; SE = standard error.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

References

Aldworth, J., Kott, P., Yu, F., Mosquin, P., & Barnett-Walker, K. (2012). Analysis of effects of 2008 NSDUH questionnaire changes: Methods to adjust adult MDE and SPD estimates and to estimate SMI in the 2005-2009 surveys. In *2010 National Survey on Drug Use and Health: Methodological resource book* (Section 16b, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.108.005). Research Triangle Park, NC: RTI International.

American Psychiatric Association. (2008). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision (DSM-IV-TR)*. Retrieved from http://www.psychiatry.org/practice/dsm/dsm-iv-tr

Center for Behavioral Health Statistics and Quality. (2009). *Results from the 2008 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2010). *Results from the 2009 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2011). *Results from the 2010 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2012a). *Results from the 2010 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2012b). *Results from the 2011 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2012c). *Results from the 2011 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2012d). *Results from the 2011 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 12-4725, NSDUH Series H-45). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2012e). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 12-4713, NSDUH Series H-44). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2013a). *Results from the 2012 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2013b). *Results from the 2012 National Survey on Drug Use and Health: Mental health detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2013c). *Results from the 2012 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 13-4805, NSDUH Series H-47). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (2013d). *Results from the 2012 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 13-4795, NSDUH Series H-46). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Center for Behavioral Health Statistics and Quality. (in press). *2012 Mental Health Surveillance Study: Design and estimation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI 0212800.001.112.008.006). Research Triangle Park, NC: RTI International.

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2012). *National Survey on Drug Use and Health: 2010 public use file codebook*. Retrieved from http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/32722

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2013). *National Survey on Drug Use and Health: 2012 public use file and codebook*. Retrieved from http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/34481

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2013). Person-level sampling weight calibration. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, Phase II, Deliverable No. 39, RTI/0211838.207.004). Research Triangle Park, NC: RTI International.

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 41, RTI/0212800.001.107.004). Research Triangle Park, NC: RTI International.

Chromy, J. R., & Penne, M. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY [CD-ROM]* (pp. 552-554). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/Proceedings/]

Dean, E., & LeBaron, P. (2009, November). *2008 National Survey on Drug Use and Health: Context effects report* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022, RTI/0209009.523.006.002). Research Triangle Park, NC: RTI International.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002, November). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research.

Frechtel, P., Carpenter, L., Edwards, S., Gulledge, K., Handley, W., King, S., Kroutil, L., Laufenberg, J., Martin, P., Moore, A., & Scott, V. (2014). Imputation report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 11, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Deliverable No. 41, RTI/0212800.001.107.006.007). Research Triangle Park, NC: RTI International.

Hughes, A., Muhuri, P., Sathe, N., & Spagnola, K. (2012). *State estimates of substance use and mental disorders from the 2009-2010 National Surveys on Drug Use and Health* (HHS Publication No. SMA 12-4703, NSDUH Series H-43). Rockville, MD: Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. [Available at http://samhsa.gov/data/NSDUH.aspx]

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

Office of Applied Studies. (2009a). *Results from the 2008 National Survey on Drug Use and Health: Detailed tables*. Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Office of Applied Studies. (2009b). *Results from the 2008 National Survey on Drug Use and Health: National findings* (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Rockville, MD: Substance Abuse and Mental Health Services Administration. [Available at http://www.samhsa.gov/data/NSDUH.aspx]

Payton, M. E., Greenstone, M. H., & Schenker, N. (2003). Overlapping confidence intervals or standard error intervals: What do they mean in terms of statistical significance? *Journal of Insect Science, 3*, 34.

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*(2), 110-123.

RTI International. (2008). *SUDAAN® language manual, Release 10.0*. Research Triangle Park, NC: RTI International.

Ruppenkamp, J., Emrich, S., Aldworth, J., Hirsch, E., & Foster, M. (2006, February). *Missingness evaluation in the 2004 NSDUH* (draft report, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-03-9028, RTI/0208726.187.022). Research Triangle Park, NC: RTI International.

Schenker, N., & Gentleman, J. F. (2001). On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician, 55*(3), 182-186.

Appendix A: Documentation for Conducting Various Statistical Procedures: SAS and SUDAAN Examples

This appendix contains example SAS® and SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2008) code that produces estimates of past month alcohol use by year (2011 and 2012) and gender (males and females) using the statistical procedures documented within this report and implemented in the 2012 detailed tables (Center for Behavioral Health Statistics and Quality [CBHSQ], 2013a) and the 2012 mental health detailed tables (CBHSQ, 2013b). The first SUDAAN example (Exhibit A.1) produces estimates as described in Section 3 of the report, and the second (Exhibit A.4) and third (Exhibit A.6) SUDAAN examples perform statistical tests of differences for the generated estimates as described in Section 7 of the report. The first SAS example (Exhibit A.2) calculates the standard error (SE) of the total for controlled domains as discussed in Section 5, and the second SAS example (Exhibit A.3) implements the suppression rule as shown in Section 10. Note that the SE of the total for uncontrolled domains is calculated within SUDAAN. The third SAS example (Exhibit A.5) produces the *p* value for the test of differences between totals for uncontrolled domains, whereas the next three SAS examples (Exhibits A.7, A.8, and A.9) combined produce the *p* value for the test of differences between totals for controlled domains. Additionally, this appendix contains example SUDAAN code that produces estimates of mean age of first use of marijuana by gender (Exhibit A.10). This example demonstrates how missing or out-of-scope values can be accounted for without subsetting the data. Exhibit A.11 demonstrates how to calculate confidence intervals. At times, a user may be interested in percentages of variable levels instead of means of variables overall; thus, Exhibit A.12 shows SUDAAN code for calculating percentages and their associated standard errors. The last two exhibits contain additional SAS and SUDAAN code related to tests of differences. Exhibit A.13 demonstrates how to test differences between two groups when the groups overlap. Exhibit A.14 demonstrates how to test independence of the prevalence variable and the subgroup variable.

Before running the SUDAAN procedures, the input dataset must be sorted by the nesting variables (VESTR and VEREP), or the NOTSORTED option must be used for SUDAAN to create an internal copy of the input dataset properly sorted by the nesting variables. The SUDAAN procedure DESCRIPT can then be run to produce weighted and unweighted sample sizes, means, totals, SEs of means and totals, and *p* values for testing of the means and totals.

The following options are specified within the SUDAAN examples to correctly produce estimates using the National Survey on Drug Use and Health (NSDUH) data.

DESIGN=WR (With Replacement)

Due to the NSDUH sample design, estimates are calculated using a method in SUDAAN that is unbiased for linear statistics. This method is based on multistage clustered sample designs where the first-stage (primary) sampling units are drawn with replacement.

Nesting Variables (VESTR and VEREP)

The nesting variables are used to capture explicit stratification and to identify clustering with the NSDUH data, which are needed to compute the variance estimates correctly. Two replicates per year were defined within each variance stratum (VESTR). Each variance replicate (VEREP) consists of four segments, one for each quarter of data collection. One replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The

other replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. A segment stays in the same VEREP for the 2 years it is in the sample. This simplifies computing SEs for estimates based on combined data from adjacent survey years.

Degrees of Freedom (DDF)

As described in Section 6 of this report, the degrees of freedom (*df*) are 900 for national estimates, 192 for large States (California, Florida, Illinois, Michigan, New York, Pennsylvania, Ohio, Texas), and 48 for all other States. For an analysis of a group of States, the degrees of freedom can be less than or equal to the sum of the degrees of freedom for each individual State due to overlap of variance strata. The specific number of degrees of freedom can be computed by counting the unique values of VESTR for the particular geographic area of interest. The technique of counting the number of unique values of VESTR can also be used for analyses combining survey data across years. When combining any years of data from 2005 through 2012, the degrees of freedom remain the same as if it were a single year (e.g., 900 for national estimates) because these years are part of the same sample design. When comparing estimates in two domains with different degrees of freedom, err on the conservative side and use the smaller degrees of freedom.

Design Effect (DEFT4)

This option within SUDAAN provides the correct measure of variance inflation due to stratification (or blocking), clustering, and unequal weighting in NSDUH estimation.

The following SAS and SUDAAN examples apply the specific NSDUH options described previously to compute estimates, apply the suppression rule, and perform significance testing by using the data produced by the example in Exhibit A.1.

Generation of Estimates

Exhibit A.1 demonstrates how to compute various types of estimates for past month alcohol use by year and gender, including the prevalence estimate (MEAN), the SE of the mean (SEMEAN), weighted sample size (WSUM), unweighted sample size (NSUM), weighted total (TOTAL), and the SE of the totals (SETOTAL). Whether the SETOTAL is taken directly from SUDAAN depends on whether the specified domain (i.e., gender in this example) is among those forced to match their respective U.S. Census Bureau population estimates through the weight calibration process. See the following section on SEs for additional information.

Exhibit A.1 SUDAAN DESCRIPT Procedure (Estimate Generation)

```
PROC SORT DATA=DATANAME; /*SAS code to sort output dataset by
Nesting Variables*/
BY VESTR VEREP;
RUN;

PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
```

Exhibit A.1 SUDAAN DESCRIPT Procedure (Estimate Generation) (continued)

```
WEIGHT ANALWT;   /*Standard single-year, person-level analysis
weight*/

VAR ALCMON;   /*Past month alcohol analysis variable*/
SUBGROUP YEAR IRSEX;
      /*Year variable, where 2011=1 & 2012=2*/
      /*Gender variable, where male=1 & female=2*/
LEVELS 2 2;
TABLES YEAR*IRSEX; /*Gender by year*/

PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL / REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM DEFFMEAN /REPLACE
   NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
   DEFFMEANFMT=F15.10 TOTALFMT=F12.0 SETOTALFMT=F12.0
   FILENAME="OUT.SUDFILE";
TITLE "ESTIMATES OF PAST MONTH ALCOHOL BY YEAR AND GENDER";
RUN;
```

Note: The following CLASS statement could be used in place of SUBGROUP and LEVELS statements in the above example:

```
CLASS YEAR IRSEX;
```

Standard Errors

As discussed in Section 5 of the report, the SE for the mean (or proportion) comes directly out of SUDAAN in the output variable SEMEAN (Exhibit A.1). However, to compute the SE of the totals, NSDUH implements different methods depending on whether the specified domain (i.e., gender in this example) is controlled or uncontrolled for during the weighting process. If a domain is uncontrolled for (i.e., it is not one of the domains described in Table 3 in Section 5), then the SE of the total comes directly out of SUDAAN in the output variable SETOTAL. If the domain is controlled for (i.e., it is one of the domains described in Table 3), then the SE of the total is calculated outside of SUDAAN as SETOTAL (SE of controlled domain) = WSUM (weighted sample size) * SEMEAN (SE for the mean/proportion). Because gender is controlled for, the SE of the totals would not be taken directly from the Exhibit A.1 example but rather would be computed using the formula shown in Exhibit A.2.

Exhibit A.2 SAS Code (Calculation of Standard Error of Totals for Controlled Domains)

```
DATA ESTIMATE;
SET OUT.SUDFILE; /*input the output file from above SUDAAN
                  procedure*/
/*************************************************************
  Define SETOTAL for gender because it is a controlled domain.
   In the SUDAAN procedure above, IRSEX is in the subgroup
      Statement with 2 levels indicated. Therefore, values for
      0=total male & females, 1=males, and 2=females are
      automatically produced.
*************************************************************/
```

Exhibit A.2 SAS Code (Calculation of Standard Error of Totals for Controlled Domains) (continued)

```
IF IRSEX IN (0,1,2) THEN SETOTAL=WSUM*SEMEAN;

RUN;
```

Suppression Rule

As described in Section 10 of the report, each published NSDUH estimate goes through a suppression rule to detect if the estimate is unreliable due to an unacceptably large sampling error. The suppression rules as they apply to different types of estimates are shown in Table 4 in Section 10. The example in Exhibit A.3 shows both the prevalence rate rule and the rule for means not bounded by 0 and 1 (i.e., averages). The average suppression rule is commented out for this example, but it would replace the prevalence rate suppression rule if averages were shown in the example in place of means bounded by 0 and 1.

For tables that display totals along with multiple means from differing populations (e.g., incidence tables in Section 4 of the 2012 detailed tables [CBHSQ, 2013a]), suppression is not as straightforward as coding the rule in the SAS/SUDAAN program. As discussed in Section 10, the case could be that some means are suppressed and others are not suppressed. In that instance, suppression of the total estimate is based on the level of suppression present across all corresponding mean estimates. If all mean estimates associated with a total estimate are suppressed, the total estimate should also be suppressed. If at least one mean estimate is not suppressed, the total estimate is also not suppressed. The best way to ensure that this happens is to program the total estimate in the table to be the one associated with the mean with the largest denominator. The analyst should also check the table after programming and table population are complete to ensure that suppression follows this rule.

Exhibit A.3 SAS Code (Implementation of Suppression Rule)

```
DATA ESTIMATE;
SET OUT.SUDFILE; /*input the output file from above SUDAAN
                  procedure*/

/******APPLY THE PREVALENCE RATE SUPRESSION RULE*******/

/* CALCULATE THE RELATIVE STANDARD ERROR */
      IF MEAN GT 0.0 THEN RSE=SEMEAN/MEAN;

/* CALCULATE THE RELATIVE STANDARD ERROR OF NATURAL LOG P */
            IF 0.0 LT MEAN LE 0.5 THEN RSELNP=RSE/ABS(LOG(MEAN));
      ELSE IF 0.5 LT MEAN LT 1.0 THEN
      RSELNP=RSE*(MEAN/(1-MEAN))/(ABS(LOG(1-MEAN)));

/*CALCULATE THE EFFECTIVE SAMPLE SIZE*/
      EFFNSUM=NSUM/DEFFMEAN;
```

Exhibit A.3 SAS Code (Implementation of Suppression Rule) (continued)

```
/*SUPRESSION RULE FOR PREVALENCE RATES*/
IF (MEAN LT .00005) OR (MEAN GE 0.99995) OR (RSELNP GT 0.175) OR
(EFFNSUM < 68) OR (NSUM <100) THEN SUPRULE=1;

/*SUPRESSION RULE FOR MEANS NOT BOUNDED BY 0 AND 1, I.E. AVERAGES
(COMMENTED OUT FOR THIS EXAMPLE)*/
/*IF (RSELNP GT 0.5) OR (NSUM < 10) THEN SUPRULE=1;*/

RUN;
```

Statistical Tests of Differences

As described in Section 7 of the report, significance tests were conducted on differences of prevalence estimates between the 2012 NSDUH and previous years of NSDUH back to 2002, as well as differences of prevalence estimates between combined 2009–2010 survey data and combined 2011–2012 survey data. Note that for year-to-year tests of differences, if the estimate for either year is suppressed, the resulting *p* value is also suppressed.

Testing of differences requires a separate PROC DESCRIPT run from the initial DESCRIPT run that produces the corresponding yearly estimates. Tests of differences can be generated using DESCRIPT's CONTRAST, PAIRWISE, or DIFFVAR statements. The SUDAAN example in Exhibit A.4 uses the DIFFVAR statement to test for differences between the 2011 and 2012 past month alcohol use estimates for all persons aged 12 or older (IRSEX=0), all males (IRSEX=1), and all females (IRSEX=2).

Similar to computing the SEs of the totals, calculating *p* values for tests of differences of totals differs depending on whether an estimate is considered to be from a controlled domain or an uncontrolled domain. Both ways are described as follows with accompanying example code: Exhibits A.4 and A.5 show example code for uncontrolled domains, and Exhibits A.4, A.6, A.7, A.8, and A.9 show example code for controlled domains.

Exhibit A.4 SUDAAN DESCRIPT Procedure (Tests of Differences)

```
PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON;
SUBGROUP YEAR IRSEX;
LEVELS 2 2;
TABLES IRSEX;
DIFFVAR YEAR=(1 2); /*Tests of differences between 2011(year=1)
                      and 2012 (year=2)*/
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
   REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
    REPLACE
    NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
    TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS”;
```

Exhibit A.4 SUDAAN DESCRIPT Procedure (Tests of Differences) (continued)

```
TITLE "TESTS OF DIFFERENCES BETWEEN 2011 AND 2012 ESTIMATES OF
PAST MONTH ALCOHOL BY YEAR AND GENDER";
RUN;
```

```
Note: The following CLASS statement could be used in place of SUBGROUP
and LEVELS statements in the above example:
      CLASS YEAR IRSEX;
```

When one or more contrasts are specified in SUDAAN, as in the DIFFVAR statement above, the output variable MEAN becomes the contrast mean, and SEMEAN becomes the SE of the contrast mean. The example above also outputs the *t*-statistic (T_MEAN) and the corresponding *p* value (P_MEAN).

SUDAAN does not test differences in the corresponding totals explicitly. However, it will output the contrast total (TOTAL) and the SE of the contrast total (SETOTAL). With these statistics and the correct degrees of freedom (900 in this example), the *p* value (PVALT) for the test of differences between totals for uncontrolled domains can be calculated as indicated in Exhibit A.5. The SAS function PROBT returns the probability from a *t*-distribution.

Exhibit A.5 SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Uncontrolled Domains)

```
IF SETOTAL GT 0.0 THEN DO;
   PVALT=2*(1-PROBT(ABS(TOTAL/SETOTAL),900));
END;
```

In the Exhibit A.1 example, all persons (aged 12 or older) and gender are annually controlled totals. For controlled domains like these, additional steps are needed to compute similar *p* values for tests of differences. One approach uses an additional DESCRIPT procedure in SUDAAN to output the appropriate covariance matrix (Exhibit A.6). Then, through further SAS data manipulations, the weighted sample sizes (WSUM), variances, and the covariance of the two means (obtained from the covariance matrix) are used to generate the standard *t*-test statistic. The corresponding *p* value can once again be produced using the SAS PROBT function and calculated *t*-test statistic.

Exhibit A.6 SUDAAN DESCRIPT Procedure (Covariance Matrix)

```
PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;
VAR ALCMON;
SUBGROUP YEAR IRSEX
LEVELS 2 2;
TABLES IRSEX*YEAR;
PRINT COVMEAN / STYLE = NCHS;
OUTPUT / MEANCOV = DEFAULT REPLACE FILENAME="OUT.SUDCOV";
TITLE "Variance Covariance Matrices ";
RUN;
```

Exhibit A.6 SUDAAN DESCRIPT Procedure (Covariance Matrix) (continued)

```
Note: The following CLASS statement could be used in place of SUBGROUP
and LEVELS statements in the above example:
      CLASS YEAR IRSEX;
```

The covariances of the estimated means can be obtained from the output of the DESCRIPT procedure (Exhibit A.6). The covariance matrix consists of a row and column for each gender (total, male, female) and year (both years, 2011, and 2012) combination with each cell corresponding to a particular variance component (i.e., a 9 x 9 matrix). Because the rows and columns of the matrix are identical, the cells in the top half (above the diagonal) and the bottom half (below the diagonal) are identical. Below is a shell of what the covariance matrix would look like for this example.

			IRSEX=0			IRSEX=1			IRSEX=2		
			YEAR=0	YEAR=1	YEAR=2	YEAR=0	YEAR=1	YEAR=2	YEAR=0	YEAR=1	YEAR=2
		ROWNUM	B01	B02	B03	B04	B05	B06	B07	B08	B09
	YEAR=0	1									
	YEAR=1	2									
IRSEX=0	YEAR=2	3									
	YEAR=0	4									
	YEAR=1	5									
IRSEX=1	YEAR=2	6									
	YEAR=0	7									
	YEAR=1	8									
IRSEX=2	YEAR=2	9									

In the SUDAAN output, each cell of the variance-covariance matrix is identified by a separate variable of the form B0*x*, where *x* is a particular cell number. (Cells are numbered left to right.) The variable *ROWNUM* is an additional output variable that simply identifies the matrix row. The covariance data needed for a particular significance test can be pulled out of the matrix using SAS code. For this example, the covariance for IRSEX=0 between YEAR=1 and YEAR=2, would be either B03 from ROWNUM2 or B02 from ROWNUM3. These two values would be the same in this case. The needed covariances are kept in the SAS code shown in Exhibit A.7.

The three SAS datasets created by the following examples, one containing the covariances (Exhibit A.7) and two containing the variances (Exhibit A.8), are then merged with the output dataset from the DESCRIPT procedure that generated the tests of differences (Exhibit A.4). With the proper statistics contained in one dataset, the corresponding *p* value for the tests of differences between controlled totals can be produced using the SAS PROBT function and calculated *t*-test statistic (Exhibit A.9).

Exhibit A.7 SAS Code (Identification of Covariance Components)

```
DATA COV(KEEP=IRSEX COV1);
  SET OUT.SUDCOV;
    IF ROWNUM=2 THEN DO; IRSEX=0; COV1=B03; END;
  ELSE IF ROWNUM=8 THEN DO; IRSEX=2; COV1=B09; END;
  ELSE IF ROWNUM=5 THEN DO; IRSEX=1; COV1=B06; END;
```

Exhibit A.7 SAS Code (Identification of Covariance Components) (continued)

```
    IF ROWNUM IN (2,5,8) THEN OUTPUT;

RUN;

PROC SORT DATA=COV; BY IRSEX; RUN;
```

The variances of the means are calculated in separate data steps shown in Exhibit A.8. The variance is simply the SE of the mean squared. The SE of the means were output in the original DESCRIPT procedure that generated the estimates.

Exhibit A.8 SAS Code (Calculation of Variances)

```
DATA EST1(KEEP=WSUM1 VAR1 YEAR IRSEX);
 SET OUT.SUDFILE;
 WHERE YEAR=1;
 WSUM1=WSUM;
 VAR1=SEMEAN**2; /*THE variance is the SEMEAN squared*/
 RUN;

 DATA EST2(KEEP=WSUM2 VAR2 YEAR IRSEX);
 SET OUT.SUDFILE;
 WHERE YEAR=2;
 WSUM2=WSUM;

RUN;
```

Exhibit A.9 SAS Code (Calculation of the *P* Value for the Test of Differences between Totals for Controlled Domains)

```
DATA P_VALUE;
 MERGE EST1 EST2 OUT.SUDTESTS COV;
 BY IRSEX;

  PVALT=2*(1-PROBT(ABS(TOTAL/SQRT(WSUM1**2*VAR1+WSUM2**2*VAR2-
        2*WSUM1*WSUM2*COV1)),900));

RUN;
```

In the example in Exhibit A.10, the mean age of first use of marijuana will be calculated in two ways. Respondents who have never used marijuana are assigned IRMJAGE=991, and if this level is included in the analysis, then the mean age calculated will be too high. Thus, two methods are shown on how to omit this level in the calculation of mean age of first use of marijuana using SAS and SUDAAN.

Exhibit A.10 SAS Code (Recoding a Variable) and SUDAAN DESCRIPT Procedure (Estimate Generation with (a) Missing Values, (b) Using Subpopulation)

/* Method 1, recoding unused values to missing*/

```
DATA DATANAME;
 SET DATANAME;
 IF IRJMAGE=991 THEN IRMJAGE_R=.;
 ELSE IRMJAGE_R=IRMJAGE;
 RUN;

 PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR
 FILETYPE=SAS DEFT4;
 NEST VESTR VEREP;
 WEIGHT ANALWT;   /*Standard single-year, person-level
 analysis weight*/
 VAR IRMJAGE_R;   /*Marijuana Age of First Use analysis
 variable*/
 SUBGROUP IRSEX;
 /*Gender variable, where male=1 & female=2*/
 LEVELS 2;
 TABLES IRSEX; /*Gender*/
 PRINT MEAN SEMEAN / REPLACE STYLE=NCHS;
 TITLE "ESTIMATES OF AGE OF FIRST USE OF MARIJUANA BY
 GENDER";
RUN;
```

/* Method 2, using subpopulation to omit the unused values*/

```
PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
NEST VESTR VEREP;
WEIGHT ANALWT;   /*Standard single-year, person-level analysis
weight*/
SUBPOPN MRJFLAG=1 /*Sub setting to omit those respondents who had
never used marijuana, i.e., omitting respondents where
IRMJAGE=991*/
VAR IRMJAGE;   /*Marijuana Age of First Use analysis variable*/
SUBGROUP IRSEX;
      /*Gender variable, where male=1 & female=2*/
LEVELS 2;
TABLES IRSEX; /*Gender*/
PRINT MEAN SEMEAN / REPLACE STYLE=NCHS;
TITLE "ESTIMATES OF AGE OF FIRST USE OF MARIJUANA BY GENDER";
RUN;
```

Confidence Intervals

As discussed in Section 8 of the report, confidence intervals can be calculated using means (MEAN) and standard errors (SEMEAN) from PROC DESCRIPT in SUDAAN. After the means and standard errors are obtained (see Exhibit A.1), the code in Exhibit A.11 can be used to create the 95 percent confidence intervals.

Exhibit A.11 SAS Code (Calculating a 95 Percent Confidence Interval for a Mean)

```
DATA CI;
SET OUT.SUDFILE; /*output data from Exhibit A.1*/
 T_QNTILE=TINV(0.975,DOF); /*define t-statistic*/

 NUMBER=SEMEAN/(MEAN*(1-MEAN));
 L=LOG(MEAN/(1-MEAN));

 A=L-T_QNTILE*NUMBER;
 B=L+T_QNTILE*NUMBER;

 PLOWER=1/(1+EXP(-A));
 PUPPER=1/(1+EXP(-B));
 /*PLOWER AND PUPPER ARE THE 95% CIS ASSOCIATED WITH MEAN
 FROM SUDAAN*/

 RUN;
```

Calculating Percentages for Categories

Exhibit A.12 demonstrates how to compute estimates corresponding to levels of a categorical variable. This example uses the number of days used marijuana in the past month among past month marijuana users. The variable that will be analyzed (MRJDAYS) is a categorical variable with days grouped into four levels (1=1-2 days, 2=3-5 days, 3=6-19 days, 4=20+ days). Because SUDAAN now needs to estimate percentages and SEs for each level of the variable instead of one estimate for the variable overall, the CATLEVEL statement is introduced and the PERCENT and SEPERCENT keywords replace the MEAN and SEMEAN keywords. Note that the suppression rule for percentages is same as the suppression rule for means given in Exhibit A.3 except PERCENT and SEPERCENT have to be divided by 100 (and thus are equivalent to MEAN and SEMEAN in the formulas).

Exhibit A.12 SAS Code (Frequency of Use, i.e., Number of Days Used Substance in the Past Month among Past Month Users)

```
PROC DESCRIPT DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
   NEST VESTR VEREP;
   WEIGHT ANALWT;   /*Standard single-year, person-level analysis
   weight*/
   VAR MRJMDAYS MRJMDAYS MRJMDAYS MRJMDAYS;   /*Marijuana Use
   frequency in the past month variable: 1=1-2 days, 2=3-5 days,
   3=6-19 days, 4=20+ days, 5=did not use in the past month*/
   CATLEVEL 1 2 3 4;  /*levels of MRJMDAYS to be shown in table*/
   SUBGROUP MRJMON;
   /*Past month marijuana use variable, where Used in past month=1 &
   did not use in past month=0*/
   LEVELS 1;
   TABLES MRJMON; /*Tables will percents among marijuana users*/
```

Exhibit A.12 SAS Code (Frequency of Use, i.e., Number of Days Used Substance in the Past Month among Past Month Users) (continued)

```
    PRINT WSUM NSUM PERCERNT SEPERCENT TOTAL SETOTAL / REPLACE
    STYLE=NCHS;
    OUTPUT WSUM PERCENT SEPERCENT TOTAL SETOTAL NSUM DEFFMEAN
    /REPLACE FILENAME="OUT.SUDFILE_FREQ";
    TITLE "FREQUENCY OF MARIJUANA USE BY PAST MONTH MARIJUANA
  USERS";RUN;
```

In addition to testing between year differences shown in Exhibit A.4, Exhibit A.13 demonstrates testing between two overlapping domains. Specifically, this exhibit shows how to use a stacked dataset to test whether past month cigarette use among the full population aged 18 or older is different from cigarette use among persons aged 18 or older who are employed full time.

Testing Between Overlapping Domains

Exhibit A.13 SAS Code (Test of Difference when Two Groups Overlap Using Stacked Data)

This code will apply when one domain is completely contained in another or when there is only partial overlap. The example below uses two domains where one is completely contained in the other. Note that the correlations between the two estimates are accounted for in this test (i.e., correlation between past month cigarette use among persons aged 18 or older and past month cigarette use among persons aged 18 or older employed full time).

```
    DATA STACKED;
    SET DATANAME(IN=A) DATANAME(IN=B); /*reading in data twice*/
    IF A THEN DO;
          INDIC=1;
          IF EMPSTAT4 IN (1,2,3,4) THEN EMPLOY=1;
          ELSE EMPLOY=0;
    END;
    ELSE IF B THEN DO;
          INDIC=2;
          IF EMPSTAT4=1 THEN EMPLOY=1;
          ELSE EMPLOY=0;
    END;
    /*create an indicator variable for the stacked data, this will be
    used in the diffvar statement in PROC DESCRIPT
    When indic=1, employ=1 represents the full population
    When indic=2, employ=1 represents those employed full time*/
    RUN;

    PROC SORT DATA=STACKED;
    BY VESTR VEREP;
    RUN;

    PROC DESCRIPT DATA=STACKED DDF=900 DESIGN=WR FILETYPE=SAS;
    NEST VESTR VEREP;
    WEIGHT ANALWT;
```

Exhibit A.13 SAS Code (Test of Difference when Two Groups Overlap Using Stacked Data) (continued)

```
VAR CIGMON;
SUBGROUP INDIC;
LEVELS 2
DIFFVAR INDIC=(1 2); /*Since we are subsetting in the next line
to employ=1, this is testing all persons 18+ vs. employed persons
18+*/
SUBPOPN CATAG18=1 AND EMPLOY=1;
PRINT WSUM NSUM MEAN SEMEAN TOTAL SETOTAL T_MEAN P_MEAN /
   REPLACE STYLE=NCHS;
OUTPUT WSUM MEAN SEMEAN TOTAL SETOTAL NSUM T_MEAN P_MEAN /
    REPLACE
    NSUMFMT=F8.0 WSUMFMT=F12.0 MEANFMT=F15.10 SEMEANFMT=F15.10
    TOTALFMT=F12.0 SETOTALFMT=F12.0 FILENAME="OUT.SUDTESTS”;
TITLE "TESTS OF DIFFERENCES BETWEEN ALL PERSONS 18 OR OLDER AND
EMPLOYED PERSONS 18 OR OLDER";
RUN;
```

Testing Independence of Two Variables when One Variable Has Three or More Levels

When comparing population subgroups defined by three or more levels of a categorical variable, log-linear chi-square tests of independence of the subgroup and the prevalence variables are conducted first to control the error level for multiple comparisons (i.e., if the goal is to compare cigarette use among several levels of employment, first test whether cigarette use is associated with employment). If Shah's Wald *F* test (transformed from the standard Wald chi-square) indicated overall significant differences, the significance of each particular pairwise comparison of interest can be tested using SUDAAN (using code similar to Exhibit A.13). Exhibit A.14 shows the code for calculating the Wald *F* test to determine whether cigarette use is associated with employment status.

Exhibit A.14 SAS Code (Test for Independence Based on a Log-Linear Model)

```
PROC CROSSTAB DATA=DATANAME DDF=900 DESIGN=WR FILETYPE=SAS DEFT4;
   NEST VESTR VEREP;
   WEIGHT ANALWT;
   CLASS CIGMON;
   SUBGROUP EMPSTAT4; /*four level employment status variable*/
   LEVELS   4;
   SETENV DECWIDTH=6 COLWIDTH=17;
   TABLES   EMPSTAT4*CIGMON;
   TEST LLCHISQ / WALDF;   /*log linear hypothesis test, wald F
   test statistic, if test statistic is significant, then reject
   null hypothesis of no interaction*/
   SETENV DECWIDTH=4 COLWIDTH=15;
   PRINT  NSUM WSUM TOTPER ROWPER COLPER STESTVAL SPVAL SDF /
          REPLACE STYLE=NCHS;
   OUTPUT  STESTVAL SPVAL SDF / REPLACE FILENAME="TEST_CHI";
  RUN;
```

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLING ERROR REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.108.005
Phase I, Deliverable No. 18

Project Director: Thomas G. Virag

Authors:

Harper Gordek
Ralph Folsom

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

July 2013

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

SAMPLING ERROR REPORT

Prepared for the 2012 Methodological Resource Book

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.108.005
Phase I, Deliverable No. 18

Authors:

Harper Gordek
Ralph Folsom

Project Director:

Thomas G. Virag

Prepared for:

Substance Abuse and Mental Health Services Administration
Rockville, Maryland 20857

Prepared by:

RTI International
Research Triangle Park, North Carolina 27709

July 2013

Acknowledgments

This report was prepared for the Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration (SAMHSA), under Contract No. HHSS283201000003C, by staff at RTI International, a trade name of Research Triangle Institute. In addition to the listed authors, significant contributors included Art Hughes of SAMHSA and Katherine B. Morton, James R. Chromy, Claudia Clark, and Margaret Smith of RTI.

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Chapter **Page**

1. Introduction 1
2. Overview of the 2012 Sample Design 3
   - 2.1 Target Population 3
   - 2.2 Design Overview 3
     - 2.2.1 5-Year Design 3
     - 2.2.2 Sample Selection at Third and Fourth Stages 6
     - 2.2.3 Creation of Variance Estimation Strata and Replicates 7
3. Computation of Relative Standard Errors and Design Effects 9
   - 3.1 Derivation of the *RSE* [ln ($\hat{p}$)] Approximation 11
4. Comparison of Observed Precision with Expected Precision 13
   - 4.1 Precision Requirements 14
   - 4.2 Observed versus Expected Precision 14
5. Comparison of Median and Mean Design Effects 19
6. Use of Domain-Specific Design Effects for Approximating Standard Error 21
7. Conclusions 31

References 33

List of Tables

Table **Page**

4.1 Observed Precision Compared with Targeted and Projected Precision, by Age Group: 2012 16

4.2 Observed Precision Compared with Projected Precision, by Age Group and Variable: 2012 17

5.1 Comparison of Median and Mean Design Effects of 56 Outcomes: 2012 20

6.1 Median Design Effects of Lifetime Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012 22

6.2 Median Design Effects of Past Year and Past Month Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012 24

6.3 Median Design Effects of Lifetime, Past Year, and Past Month Licit Drug Use Estimates, by Age Group, Gender, and Demographic Characteristic: 2012 26

6.4 Design Effects, by Age Group, for the Outcomes Used in the Medians in Tables 6.1, 6.2, and 6.3: 2012 28

1. Introduction

As part of any survey data analysis, a good understanding of the resulting standard errors (SEs) and design effects (DEFFs), corresponding to a key set of outcomes and other variables, is important for a number of reasons: (1) to evaluate how well the sample was designed in light of the target and realized precisions and DEFFs; (2) to obtain confidence intervals (CIs) for cross-sectional estimates (and for change estimates in the case of repeated surveys) and generate significance tests to make inferential statements about the target population; and (3) to be able to incorporate realized DEFFs for future redesigns of the survey.

This report compares the estimated (or realized) precisions of a key set of estimates with the targets for the 2012 National Survey on Drug Use and Health (NSDUH), formerly called the National Household Survey on Drug Abuse (NHSDA). The comparison was made with targets specified by the Substance Abuse and Mental Health Services Administration (SAMHSA) and with the predicted precisions that statisticians from RTI International[1] anticipated during the design of the survey. In addition, tables of realized DEFFs are provided.

This report is organized as follows. Chapter 2 summarizes the 2012 sample design. Chapter 3 describes the calculation of relative standard errors (RSEs) and DEFFs. Chapter 4 presents tables that compare the observed precisions with the expected precisions. Chapter 5 compares median and mean DEFFs. Chapter 6 presents median and mean DEFFs for specific analysis domains. Finally, concluding remarks are provided in Chapter 7.

[1] RTI International is a trade name of Research Triangle Institute.

2. Overview of the 2012 Sample Design

2.1 Target Population

The respondent universe for the 2012 National Survey on Drug Use and Health[2] (NSDUH) was the civilian, noninstitutionalized population aged 12 years or older residing in the United States. Consistent with the NSDUH designs since 1991, the 2012 NSDUH universe included residents of noninstitutional group quarters (e.g., shelters, rooming houses, dormitories, and group homes), residents of Alaska and Hawaii, and civilians residing on military bases. Persons excluded from the 2012 universe included those with no fixed household address (e.g., homeless transients not in shelters) and residents of institutional group quarters, such as jails and hospitals.

2.2 Design Overview

The Substance Abuse and Mental Health Services Administration (SAMHSA) implemented major changes in the way NSDUH would be conducted, beginning in 1999 and continuing through subsequent years. The survey is conducted using computer-assisted interviewing (CAI) methods and provides State estimates based on minimum sample sizes per State. The total targeted sample size of 67,500 is equally allocated across three age groups: persons aged 12 to 17, persons aged 18 to 25, and persons aged 26 or older. This large sample size allows SAMHSA to report more precise demographic subgroup estimates at the national level without needing to oversample specially targeted demographics (such as racial/ethnic groups), as required in the past. This large sample is referred to as the "main sample." The achieved sample for the 2012 sample was 68,309 persons.

Beginning with the 2002 NSDUH and continuing through the 2012 NSDUH, survey respondents were given a $30 incentive payment for participation. As expected, the incentive had the effect of increasing response rates, thereby requiring fewer selected households than previous surveys. In recent years, however, response rates have been slowly declining, which has required the number of selected households to increase.

An additional design change was made in 2002 and continued in 2012. A new pair sampling strategy was implemented that increased the number of pairs selected in dwelling units (DUs) with older persons rostered (Chromy & Penne, 2002). With the increase in the number of pairs came a moderate decrease in the response rate for older persons.

2.2.1 5-Year Design

A coordinated sample design was developed for the 2005 through 2009 NSDUHs. The 2010 through 2012 samples are extensions of the 5-year sample. Although there was no planned overlap with the 1999 through 2004 samples, a coordinated design for 2005 through 2009 facilitated 50 percent overlap in second-stage units (area segments) within each successive 2-

[2] This report presents information from the 2012 National Survey on Drug Use and Health (NSDUH). Prior to 2002, the survey was called the National Household Survey on Drug Abuse (NHSDA).

year period from 2005 through 2009. The intent of this design feature is to increase the precision of estimates in year-to-year trend analyses, using the expected positive correlation resulting from the overlapping sample between successive NSDUH years. The 2012 NSDUH main sample continues the 50 percent overlap by retaining half of the second-stage units from the 2011 survey.

The 2012 design provided for estimates by State in all 50 States plus the District of Columbia. States may, therefore, be viewed as the first level of stratification as well as a reporting variable. Eight States, referred to as the "large" States,[3] had samples designed to yield 3,600 respondents per State for the 2012 survey. This sample size was considered adequate to support annual direct State estimates. The remaining 43 States[4] had samples designed to yield 900 respondents per State in the 2012 survey. In these 43 States, adequate data were available to support reliable State estimates based on small area estimation (SAE) methodology. Reliable direct State estimates are also possible (in any State) by pooling multiple years of data.

State sampling (SS) regions were formed within each State. Based on a composite size measure, States were geographically partitioned into regions of roughly equal size. In other words, regions were formed such that each area yielded, in expectation, roughly the same number of interviews during each data collection period, thus distributing the workload equally among NSDUH interviewers. The smaller States were partitioned into 12 SS regions, whereas the eight large States were divided into 48 SS regions. Therefore, the partitioning of the United States resulted in the formation of a total of 900 SS regions.

Unlike the 1999 through 2001 NHSDAs and the 2002 through 2004 NSDUHs, the first stage of selection for the 2005 through 2012 NSDUHs was census tracts.[5] This stage was included to contain sample segments within a single census tract to the extent possible.[6] In prior years, segments that crossed census tract boundaries made merging to external data sources difficult.

The first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated within SS regions until each tract[7] had, at a minimum, 150 DUs[8] in urban areas and 100 DUs in rural areas.[9]

---

[3] The large States are California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas.

[4] For reporting and stratification purposes, the District of Columbia is treated the same as a State, and no distinction is made in the discussion.

[5] A census tract is a small, relatively permanent statistical subdivision of a county or equivalent entity that contains between 1,200 and 8,000 people, with an optimum size of 4,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

[6] Some census tracts had to be aggregated in order to meet the minimum DU requirement.

[7] For the remainder of the discussion, first-stage sampling units will be referred to as "census tracts" even though each first-stage sampling unit contains one or more census tracts.

[8] DU counts were obtained from the 2000 census data supplemented with revised population counts from Nielsen Claritas.

[9] The basis for the differing minimum DU requirement in urban and rural areas is that it is more difficult to meet the requirement in rural areas, and 100 DUs is sufficient to support one field test and two main study samples.

Before selecting census tracts, additional implicit stratification was achieved by sorting these first-stage sampling units by a CBSA/SES[10] (core-based statistical area/socioeconomic status) indicator[11] and by the percentage of the population that is non-Hispanic or Latino and white.[12] From this well-ordered sample frame, 48 census tracts per SS region were selected with probabilities proportionate to a composite size measure and with minimum replacement (Chromy, 1979).

Because census tracts generally exceed the minimum DU requirement, one smaller geographic region was selected within each sampled census tract. For this second stage of sampling, each selected census tract was partitioned into compact clusters[13] of DUs by aggregating adjacent census blocks.[14] Consistent with the terminology used in previous NSDUHs, these geographic clusters of blocks are referred to as "segments." A sample DU in NSDUH refers to either a housing unit or a group-quarters listing unit, such as a dormitory room or a shelter bed. Similar to census tracts, segments were formed to contain a minimum of 150 DUs in urban areas and 100 DUs in rural areas. This minimum DU requirement supports the overlapping sample design and any special supplemental samples or field tests that SAMHSA may wish to conduct.

Prior to selection, the segments were sorted in the order they were formed (i.e., geographically), and one segment was selected within each sampled census tract using Chromy's method of sequential random sampling (with probability proportionate to size and minimum replacement) (Chromy, 1979). The 48 selected segments then were randomly assigned to a survey year and quarter of data collection as described in Section 2.4 of the 2012 sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013).

---

[10] CBSAs include metropolitan and micropolitan statistical areas as defined by the Office of Management and Budget (2003).

[11] Four categories are defined as (1) CBSA/low SES, (2) CBSA/high SES, (3) non-CBSA/low SES, and (4) non-CBSA/high SES. To define SES, census tract–level median rents and property values obtained from the 2000 Census Summary File 3 were given a rank (1,…,5) based on State and CBSA quintiles. The rent and value ranks then were averaged, weighted by the percentages of renter- and owner-occupied DUs, respectively. If the resulting score fell in the lower 25th percentile by State and CBSA, the area was considered "low SES"; otherwise, it was considered "high SES."

[12] Although the large sample size eliminates the need for the oversampling of specially targeted demographic subgroups as was required prior to the 1999 NHSDA, sorting by a CBSA/SES indicator and by the percentage of the population that is non-Hispanic or Latino and white ensures dispersion of the sample with respect to SES and race/ethnicity. Implicit stratification also has the potential to lower sampling error by reducing the selection of neighboring and possibly similar segments than if the selection was done completely at random.

[13] Although the entire cluster is compact, the final sample of DUs represents a noncompact cluster. Noncompact clusters (selection from a list) differ from compact clusters in that not all units within the cluster are included in the sample. Although compact cluster designs are less costly and more stable, a noncompact cluster design was used because it provides for greater heterogeneity of dwellings within the sample. Also, social interaction (contagion) among neighboring dwellings is sometimes introduced with compact clusters (Kish, 1965).

[14] A census block is a small statistical area bounded by visible features (e.g., streets, roads, streams, and railroad tracks) and nonvisible boundaries (e.g., city, town, and county limits). A block group is a cluster of census blocks within the same census tract and generally contains between 300 and 6,000 people (U.S. Census Bureau, Redistricting Data Office, 2009).

2.2.2 Sample Selection at Third and Fourth Stages

Once sample segments for the 2012 NSDUH were selected, specially trained field household listers visited the areas and obtained complete and accurate lists of all eligible DUs within the sample segment boundaries. These lists served as the frames for the third stage of sample selection.

The primary objective of the third stage of sample selection (listing units) was to determine the minimum number of DUs needed in each segment to meet the targeted sample sizes for all age groups. Thus, listing unit sample sizes for the segment were determined using the age group with the largest sampling rate, which we refer to as the "driving" age group. Using 2000 census data adjusted to more recent data from Claritas, Inc., State- and age-specific sampling rates were computed. These rates then were adjusted by the segment's probability of selection; the subsegmentation inflation factor,[15] if any; the probability of selecting a person in the age group (equal to the maximum, or 0.99, for the driving age group); and an adjustment for the "maximum of two" rule.[16] In addition to these factors, historical data from the 2010, 2011, and 2012 NSDUHs were used to compute predicted screening and interviewing response rate adjustments. The final adjusted sampling rate then was multiplied by the actual number of DUs found in the field during counting and listing activities. The product represents the segment's listing unit sample size.

Some constraints were put on the listing unit sample sizes. For example, to ensure adequate samples for supplemental studies, the listing unit sample size could not exceed 100 per segment or half of the actual listing unit count. Similarly, if five unused listing units remained in the segment, a minimum of five listing units per segment were required for cost efficiency.

Using a random start point and interval-based (systematic) selection, the actual listing units were selected from the segment frame. After DU selections were made, an interviewer visited each selected DU to obtain a roster of all persons residing in the DU. As in previous years, during the data collection period, if an interviewer encountered any new DU in a segment or found a DU that was missed during the original counting and listing activities, the new or missed dwellings were selected into the 2012 NSDUH using the half-open interval selection technique.[17] This selection technique eliminates any frame bias that might be introduced because of errors and/or omissions in the counting and listing activities, and it also eliminates any bias that might be associated with using "old" segment listings.

---

[15] Segments found to be very large in the field are partitioned into "subsegments." Then, one subsegment is chosen at random with probability proportional to the size to be fielded. In some cases, a second-level subsegmenting was required if the census totals used in the initial subsegmenting were off and the selected subsegment was still too large for listing. The subsegmentation inflation factor accounts for the narrowing down of the segment.

[16] Brewer's Selection Algorithm never allows for greater than two persons per household to be chosen. Thus, sampling rates are adjusted to satisfy this constraint.

[17] In summary, this technique states that, if a DU is selected for the 2012 study and an interviewer observes any new or missed DUs between the selected DU and the DU appearing immediately after the selection on the counting and listing form, all new or missed DUs falling in this interval will be selected. These added DUs are assigned the same probability of selection as the selected DU. If a large number of new or missed DUs are encountered (greater than 10), a sample of the new or missing DUs will be selected, and the sample weight will be adjusted accordingly.

Using the roster information obtained from an eligible member of the selected DU, 0, 1, or 2 persons were selected for the survey. Sampling rates were preset by age group and State. Roster information was entered directly into the electronic screening instrument, which automatically implemented this fourth stage of selection based on the State and age group sampling parameters.

One benefit of using an electronic screening instrument in NSDUH is the ability to impose a more complicated person-level selection algorithm on the fourth stage of the NSDUH design. Similar to the 1999 through 2011 designs, one feature that was included in the 2012 design was that any two survey-eligible persons within a DU had some chance of being selected (i.e., all survey-eligible pairs of persons had some nonzero chance of being selected). This design feature was of interest to NSDUH researchers because, for example, it allows analysts to examine how the drug use propensity of one individual in a family relates to the drug use propensity of other family members residing in the same DU (e.g., the relationship of drug use between a parent and his or her child).

2.2.3 Creation of Variance Estimation Strata and Replicates

The nature of the stratified, clustered sampling design requires that the design structure be taken into consideration when computing variances of survey estimates. Key nesting variables representing the variance estimation strata and replicates were created to capture explicit stratification and to identify clustering. For the 2005 through 2013 NSDUHs, variance estimation strata are defined at the SS region level, and each SS region is assigned to a different stratum every quarter in pseudo-random fashion. Two replicates per year were defined within each variance stratum. Each variance replicate consists of four segments, one for each quarter of data collection. The first replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The second replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years.

Both census tracts and segments are nested within variance replicates, so the nesting variables cover the variance contributions of both sampling units. Because one segment is selected per sampled census tract, the selection of census tracts at the first stage of selection may reduce variance by controlling the sample distribution and minimizing the chance of selecting neighboring and possibly similar segments within the same census tract. All weighted statistical analyses for which variance estimates are needed should use the stratum and replicate variables to identify nesting. Variance estimates can be computed using a clustered data analysis software package such as the SUrvey DAta ANalysis program (SUDAAN®) (RTI International, 2008) (Morton et al., 2013).

3. Computation of Relative Standard Errors and Design Effects

As mentioned in Chapter 1, there were several objectives for calculating relative standard errors (RSEs) and design effects (DEFFs) for the 2012 National Survey on Drug Use and Health (NSDUH). One objective was to provide a mechanism for comparing the expected precision of the 2012 design with the precision actually obtained. A second objective was to provide users of NSDUH data with a methodology for determining a quick approximation of the precision of estimates obtained from the 2012 survey. The third objective was to build confidence intervals (CIs) of estimates of level and change. Finally, the magnitudes of the DEFFs are useful for future redesign of the survey.

The RSE of a domain *d* prevalence estimate is the standard error (SE) of the estimate divided by the estimate, that is,

$$RSE(\hat{P}_d) = SE(\hat{P}_d) / \hat{P}_d. \tag{1}$$

The DEFF for a prevalence estimate is its variance divided by the variance that would be observed if simple random sampling (SRS) had been used. Hence, the SE of the estimated prevalence can be approximated as follows:

$$SE(\hat{P}_d) \doteq \left[DEFF(d)\hat{P}_d(1-\hat{P}_d)/n_d\right]^{1/2}, \tag{2}$$

where $DEFF(d)$ and n_d are the median (or mean, as the case may be) DEFF and sample size of domain *d*, respectively.

By substituting a prevalence rate of 0.10 into Equations 1 and 2, the RSE becomes

$$RSE(\hat{P}_d = 0.10) = \left[(DEFF(d) * 0.09 / n_d)\right]^{1/2}. \tag{3}$$

This result shows that, for the specified prevalence rate of 0.10, the RSE is purely a function of the DEFF and sample size. In the tables given in this report, RSEs are expressed as percentages (i.e., the right-hand side of Equation 3 is multiplied by 100).

Mean and median DEFFs were used for many of the calculations in this report. DEFFs were calculated based on drug use variables displayed in the 2012 NSDUH sample design report (Morton et al., 2013).

As noted previously, the DEFF is the ratio of the design-based variance estimate divided by the variance estimate that would have been obtained from an SRS of the same size. Therefore, the DEFF summarizes the effects of stratification, clustering, and unequal weighting on the variance of a complex sample design. Because clustering and unequal weighting are expected to increase the variance and generally dominate the stratification effect, the DEFF is expected to be

greater than 1 in most instances. However, DEFFs were sometimes less than 1 for prevalence rates near 0.

Note that the DEFF is based on the with-replacement (wr) variance estimate as obtained from SUDAAN®, which properly accounts for clustering, stratification, and unequal weighting (RTI International, 2008). In the 1999 sampling error report, DEFF was based on the maximum-of-three rule for computing design-based SEs under the premise that the precision loss anticipated due to clustering and unequal probability sampling offsets any gain due to stratification (i.e., the DEFF should be at least 1). The three SEs correspond to the SUDAAN assumption of wr primary sampling units (PSUs), stratified simple random sample, and simple random sample. Note that, for the 2000 survey onward, it was decided to use only the standard SUDAAN wr SE, based on the PSU for the sake of simpler interpretation, as well as for easier computation of the SE of functions of estimates, such as differences and ratios.

In addition, starting in 2005 and continuing into 2012, the survey has seen the full implementation of a change made to the method of calculation for the SEs of estimated totals. It had been discovered in previous survey years that the original method, multiplying the weighted sample size by the corresponding SE of the mean, had produced underestimation of the SE of some estimated totals that had not been controlled for through the weighting process. As a result, from 2000 to 2003, in order to better reflect variance of the estimates within uncontrolled domains, a subset of the detailed tables reported the SE of the estimated totals directly from the SUDAAN calculation.

However, applying the SEs directly from SUDAAN to only a subset of tables produced inconsistencies between different sets of tables. Therefore, in 2004 and continuing through 2012, the process changed so that the decision about which method of calculation would be used for the SEs of estimated totals was made at the estimate level (e.g., the cell level) rather than at the marginal table level. This way the estimated totals would have consistent values for their variances throughout all reported tables. A specific set of domains used as covariates in the poststratification step of the NSDUH weighting process were designated as the "controlled" domains. The SE reported for these domains would be based on the original method. All other domains would report the SE directly from the SUDAAN calculation. A more detailed discussion on the change in SE reporting can be found in Appendix B, Section B.2.1, of the 2012 summary of national findings (Center for Behavioral Health Statistics and Quality, in press).

DEFFs associated with prevalence estimates below 0.00005 or greater than or equal to 0.99995 (an ad hoc rule representing 0 or 1 in practice) or prevalence estimates exhibiting low precision were not used for determining the medians. To identify estimates with low precision, the suppression rule used in earlier years was applied. Specifically, DEFFs or the corresponding prevalence estimates were not included if the corresponding RSE of -1n*(p)* satisfies

$$RSE[\text{-1n}(p)] > 0.175 \text{ when } p \leq 0.5$$

or

$$RSE[\text{-1n}(1-p)] > 0.175 \text{ when } p > 0.5.$$

A rationale for this rule is that, for a prevalence estimate of 0.10, the minimum required effective sample size (or the sample size under SRS) is around 55 (55.43 to be exact) when the

maximum tolerable value of $RSE[-\ln(p)] = 0.175$. This can be derived as follows: under SRS, $RSE(p)$ is equal to the square root of $p(1-p)/np^2$, and using a first-order Taylor series, $SE[-\ln(p)]$ is approximately $SE(p)/p$ (i.e., $RSE(p)$). Therefore, under SRS, $RSE[-\ln(p)]$ is approximately $RSE(p)/[-\ln(p)]$. Then, substituting $p = 0.10$ and $RSE[-\ln(p)] = 0.175$ gives $n = 55.43$ under SRS. For complex designs where $DEFF(p) > 1$, this can be interpreted as the minimum required effective sample size. In other words, if $DEFF(p) = 2$, the minimum required sample size is the DEFF times the effective sample size (i.e., 111). The derivation of $RSE[-\ln(p)]$ is given below.

It may be noted that, for a given sample size, the RSE increases as p decreases, and for a given p, it increases as the sample size decreases. The above discussion pertains to $p < 0.5$. Although the RSE of p is not symmetric about $p = 0.5$, it makes logical sense for precision requirements to be identical for p and $1 - p$. Therefore, it is convenient to use the convention that the suppression rule for $p < 0.5$ also applies for $p > 0.5$ by replacing p with $1 - p$.

3.1 Derivation of the $RSE[\ln(\hat{p})]$ Approximation

Define the first-order Taylor series of a function, $f(\hat{\theta})$, about a point, θ, as

$$f(\hat{\theta}) \cong f(\theta) + \left[df(\hat{\theta})/d\hat{\theta}\right]_{\hat{\theta}=\theta}(\hat{\theta}-\theta),$$

then

$$E(\hat{\theta}) = \theta \text{ and } \left[f(\hat{\theta}) - f(\theta)\right] \cong \left[df(\hat{\theta})/d\hat{\theta}\right]_{\hat{\theta}=\theta}(\hat{\theta}-\theta) \text{ lead to}$$

$$Var\left[f(\hat{\theta})\right] = E\left[f(\hat{\theta}) - f(\theta)\right]^2 \cong \left[df(\hat{\theta})/d\hat{\theta}\right]^2_{\hat{\theta}=\theta} Var(\hat{\theta}), \text{ where } Var(\hat{\theta}) = E(\hat{\theta}-\theta)^2.$$

Let

$$\hat{\theta} = \hat{p} \quad f(\hat{\theta}) = \ln(\hat{p}) \quad d\ln(\hat{p})/d\hat{p} = 1/p,$$

then

$$Var[\ln(\hat{p})] \cong Var(\hat{p}) \div p^2 = [RSE(\hat{p})]^2,$$

and

$$Rel - Var\left[\ln\left(\hat{p}\right)\right] \cong \left[RSE\left(\hat{p}\right)\right]^2 \div \left[\ln\left(p\right)\right]^2.$$

This leads to the estimated $RSE\left[\ln\left(\hat{p}\right)\right]$:

$$RSE\left[\ln\left(\hat{p}\right)\right] \cong RSE\left(\hat{p}\right) \div \left[-\ln\left(\hat{p}\right)\right].$$

4. Comparison of Observed Precision with Expected Precision

In this chapter, two benchmarks in the 2012 National Survey on Drug Use and Health (NSDUH) are compared with the estimated achieved precision of important outcome measures. One benchmark is derived from requirements specified by the Substance Abuse and Mental Health Services Administration (SAMHSA), and the other is the predicted precision that statisticians at RTI International anticipated during the design of the survey.

Predicted precision requirements for the 2012 designs were specified in terms of targeted relative standard errors (RSEs) and in terms of minimum sample sizes. To obtain the targeted RSEs, historical 2001 data were used, and RSEs were computed for nine outcome measures of interest. RSEs were then standardized to a prevalence of 10 percent as given by Equation 3 in Chapter 3.

The nine key NSDUH outcomes that the sample design optimization for the 2012 NSDUH was based on included recency-of-use estimates, treatment received for alcohol and illicit drug use, and dependence on alcohol and illicit drug use. Specifically, the following outcomes were used for 2012 (variable names on the NSDUH data files are in parentheses):

- cigarette use in the past month (CIGMON),
- alcohol use in the past month (ALCMON),
- any illicit drug use in the past month (SUMMON),
- any illicit drug use other than marijuana in the past month (IEMMON),
- cocaine use in the past month (COCMON),
- dependent on illicit drugs in the past year (DEPNDILL),
- dependent on alcohol in the past year (DEPNDALC) but not dependent on illicit drugs in the past year (DEPNDILL),
- received treatment at any location for illicit drug use in the past year (ILLTRMT), and
- received treatment for alcohol at any location in the past year (ALCTRMT) but did not receive treatment for illicit drugs in the past year (ILLTRMT) at any location.

Table 4.1 at the end of this chapter shows an overall age group comparison of the benchmark and observed design effects (DEFFs) and RSEs based on the nine outcomes from the 2012 sample design report (Morton et al., 2013). Table 4.2 at the end of this chapter shows a comparison of the projected and observed DEFFs and RSEs for the nine outcomes from the sample design report by overall and age group breakdowns.

4.1 Precision Requirements

Initial requirements for the sample were set as follows:

- minimum sample sizes of 3,600 persons per State in the 8 largest States and 900 persons in the remaining 43 States; and
- equal allocation of the sample across the three age groups: 12 to 17, 18 to 25, and 26 or older within each State.

In addition, for national estimates, the SAMHSA-specified precision requirements were that the expected RSE on a prevalence of 10 percent not exceed the following:

- 3.0 percent for total population statistics; and
- 5.0 percent for statistics in three age group domains: 12 to 17, 18 to 25, and 26 or older.

The 2012 sample reflected SAMHSA's objective to develop reliable State-level estimates using small area estimation (SAE) procedures. To achieve this objective, the targeted sample size by State was set at 900 completed interviews. In the eight largest States, the target was set at 3,600 completed interviews. The larger overall sample made it possible to get adequate precision for Hispanic or Latino and non-Hispanic or Latino black or African-American populations at the national level without any targeted oversampling of areas of high concentration of these populations or any oversampling through screening for these target populations.

4.2 Observed versus Expected Precision

The tables at the end of this chapter present observed results compared with benchmarks for sample sizes, DEFFs, and associated RSEs. For Table 4.1, the projected RSEs were the averages over the nine outcome variables as given in the beginning of this chapter. Note that, using Equation 3, the RSEs for all the outcome variables were scaled to the generic prevalence of 0.10. The projected DEFF was derived as an average over the DEFFs for the nine variables corresponding to the projected RSEs via Equation 3 for various domains. For the observed RSE, as in previous years' reports, mean DEFFs for the nine outcomes were substituted into Equation 3 to obtain mean RSEs for a prevalence of 0.10. The mean was used here for comparison purposes instead of the median because the mean was used for the purpose of sample allocation. Also, because the DEFF was proportional to the squared RSE or relative variance, it is probably more meaningful to compute projected RSE over all nine outcomes as root mean relative variance rather than mean RSE. However, the difference between the two was only marginal.

All of the nine prevalence estimates contributed to the means in Table 4.1. None were suppressed because of low precision. The observed DEFFs were generally slightly higher than the projections. Specifically, only the 12–17 age group produced smaller observed statistics. However, all of the ones that were higher were within 6 percent of the projected values. In comparison with the target values, the observed RSEs fared better. They were lower than the targets by roughly half in two cases and more than 30 percent in all.

In Table 4.2, the comparison is between the observed and projected precisions for each of the nine outcomes used in sample design optimization. Although the observed mean DEFF values vary quite a bit from the projections, it is important to note that the observed mean RSE values, the values used in the sample design report, are relatively close to the projected values. Out of 36 observed values, only 3 are larger than their projections by greater than 10 percent, and more than half are below their projections.

Table 4.1 Observed Precision Compared with Targeted and Projected Precision, by Age Group: 2012

	Sample Size			Mean Design Effect			Mean Relative Standard Error at p = 10 Percent			
Age Group	**Projected**	**Observed**	**% Diff**	**Projected**	**Observed**	**% Diff**	**Projected**	**Target**	**Observed[1]**	**% Diff[2]**
Total	67,500	68,309	1.20%	2.99	3.16	5.8%	1.99	3.00	2.03	-32.3%
12–17	22,500	22,473	-0.12%	1.95	1.78	-8.9%	2.79	5.00	2.66	-46.7%
18–25	22,500	22,529	0.13%	2.06	2.09	1.5%	2.87	5.00	2.89	-42.2%
26+	22,500	23,307	3.59%	1.72	1.79	4.2%	2.62	5.00	2.63	-47.5%

[1]Calculated using Equation 2 with the observed sample size and the mean observed design effect.

[2]Percent relative difference from the target relative standard error.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 4.2 Observed Precision Compared with Projected Precision, by Age Group and Variable: 2012

Age Group	Drug	Mean Design Effect			Mean Relative Standard Error at p = 10 Percent		
		Projected	Observed	% Diff	Projected	Observed[1]	% Diff[2]
Total	CIGMON	3.63	4.12	13.6%	2.20	2.33	6.0%
	ALCMON	4.11	4.18	1.8%	2.34	2.35	0.3%
	SUMMON	3.06	2.94	-4.1%	2.02	1.97	-2.6%
	IEMMON	2.74	2.92	6.6%	1.91	1.96	2.6%
	COCMON	2.12	3.25	53.6%	1.68	2.07	23.2%
	DEPNDILL	2.35	2.44	3.8%	1.77	1.79	1.3%
	DEPNDALC[3]	3.24	2.87	-11.4%	2.08	1.95	-6.4%
	ILLTRMT	2.38	2.34	-1.4%	1.78	1.76	-1.3%
	ALCTRMT[3]	3.31	3.41	3.2%	2.10	2.12	1.0%
12–17	CIGMON	1.95	1.78	-8.7%	2.79	2.67	-4.4%
	ALCMON	2.09	1.95	-6.8%	2.89	2.79	-3.4%
	SUMMON	2.06	1.69	-17.9%	2.87	2.60	-9.4%
	IEMMON	1.97	1.90	-3.8%	2.81	2.76	-1.9%
	COCMON	1.92	1.40	-27.0%	2.77	2.37	-14.5%
	DEPNDILL	1.74	1.83	5.3%	2.64	2.71	2.7%
	DEPNDALC[3]	2.13	1.89	-11.4%	2.92	2.75	-5.8%
	ILLTRMT	1.88	1.97	5.1%	2.74	2.81	2.6%
	ALCTRMT[3]	1.82	1.58	-13.1%	2.70	2.52	-6.7%
18–25	CIGMON	2.21	2.26	2.6%	2.97	3.01	1.2%
	ALCMON	2.36	2.28	-3.4%	3.07	3.02	-1.8%
	SUMMON	2.10	2.13	1.1%	2.90	2.91	0.5%
	IEMMON	1.86	2.07	10.9%	2.73	2.87	5.3%
	COCMON	2.02	1.77	-12.4%	2.84	2.66	-6.5%
	DEPNDILL	2.18	2.16	-0.5%	2.95	2.94	-0.3%
	DEPNDALC[3]	2.02	1.88	-6.5%	2.84	2.74	-3.4%
	ILLTRMT	1.85	2.44	32.0%	2.72	3.12	14.8%
	ALCTRMT[3]	1.99	1.86	-6.4%	2.82	2.73	-3.3%
26+	CIGMON	1.88	2.11	12.6%	2.74	2.86	4.3%
	ALCMON	2.13	2.22	4.1%	2.92	2.93	0.3%
	SUMMON	1.88	1.81	-3.5%	2.74	2.64	-3.5%
	IEMMON	1.76	1.74	-0.7%	2.65	2.59	-2.1%
	COCMON	1.24	1.82	46.7%	2.23	2.65	19.0%
	DEPNDILL	1.64	1.73	5.3%	2.56	2.58	0.8%
	DEPNDALC[3]	1.84	1.58	-13.8%	2.71	2.47	-8.8%
	ILLTRMT	1.39	1.40	0.7%	2.36	2.33	-1.4%
	ALCTRMT[3]	1.74	1.73	-0.9%	2.64	2.58	-2.2%

[1]Calculated using Equation 2 with the observed sample size and the mean observed design effect.

[2]Percent relative difference from the target relative standard error.

[3]Results reported for DEPNDALC and ALCTRMT are defined by multiple variables as noted at the beginning of Chapter 4.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

5. Comparison of Median and Mean Design Effects

The mean design effect (DEFF) is more sensitive to outliers and is generally larger than the median. Table 5.1 compares the median and mean of 56 DEFFs for three age groups and over all ages in the 2012 design for the National Survey on Drug Use and Health (NSDUH). Comparisons are also provided for the four race/ethnicity categories, although they were not used as stratification variables when selecting persons within households.

The median and mean DEFF estimates were based on estimates from the following:

- *15 illicit drug use categories:* any illicit drug use, marijuana/hashish, cocaine, crack, inhalants, hallucinogens, LSD, PCP, heroin, nonmedical use of any psychotherapeutics, nonmedical use of stimulants, nonmedical use of sedatives, nonmedical use of tranquilizers, nonmedical use of pain relievers, any illicit drug except marijuana; and
- *3 licit drug use categories:* cigarettes, alcohol, and smokeless tobacco.

Estimates used from these categories included one from each of three recency-of-use classes: ever used, used in past year, and used in past month.

The estimates of past month heavy drinking and binge drinking also were included in the licit drug use category, bringing the total number of estimates used for the mean versus median comparisons to 56. The median and the mean DEFF were calculated from the above estimates for the total population, by age and by race/ethnicity. As seen from Table 5.1, the mean DEFF turned out to be larger than the median DEFF in four of the eight domains. The differences between the mean and median DEFFs fall below 5 percent for the total and age comparison groups. However, for the race/ethnicity comparison groups, the percentage differences vary by more than 7.5 percent for all non-white groups.

Table 5.1 Comparison of Median and Mean Design Effects of 56 Outcomes: 2012

Outcome	Median Design Effect	Mean Design Effect	Difference (Mean – Median)	Percentage Difference[1]
Total	3.16	3.09	-0.07	-2.17
Age (Years)				
12–17	1.69	1.77	0.08	4.49
18–25	2.11	2.08	-0.03	-1.28
26+	1.80	1.82	0.02	1.01
Race/Ethnicity				
White, Not Hispanic or Latino	2.80	2.79	-0.01	-0.47
Black or African American, Not Hispanic or Latino	3.39	3.78	0.39	11.52
Hispanic or Latino	3.68	3.32	-0.36	-9.75
Other or Multiple, Not Hispanic or Latino	2.86	3.28	0.42	14.59

[1]Computed as 100*(Mean – Median)/Median.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

6. Use of Domain-Specific Design Effects for Approximating Standard Error

This chapter presents an approach considered for approximating the standard error (SE) when published 2012 National Survey on Drug Use and Health (NSDUH) SE estimates or computer software is unavailable. The approach considered in this chapter is based on domain-specific median design effects (DEFFs).

Domains were defined by cross-classifications of age and gender, race/ethnicity, population density, geographic division of residence, adult education, current employment, and State.[18] The 56 types of drug and recency categories, provided in Chapter 5, were used for the estimates on which the medians were computed. DEFFs associated with percentage estimates exhibiting low precision, as defined in Chapter 3, were not used. The median DEFFs were computed separately for the three classifications: lifetime illicit drug use (Table 6.1), past year and past month illicit drug use (Table 6.2), and licit drug use (Table 6.3). Note that DEFFs for lifetime use were expected to be quite different from those for past year use and past month use; therefore, it was desirable to keep the two separate. However, this separation was not done for licit drugs because of the small number of drug use variables available for computing the median for each domain (a total of only 11). These tables can be used to calculate an approximate variance estimate for a particular domain as follows:

$$var\left(p_d\right)_{appx} = DEFF_{d,MED} * \left[p_d\left(1-p_d\right)/n_d\right], \tag{4}$$

where p_d is the estimated proportion for domain *d*, n_d is the sample size for domain *d*, and $DEFF_{d,MED}$ is the median design effect for domain *d*.

The approximate SE estimate for p_d, $SE(p_d)_{appx}$, is the square root of $var(p_d)appx$. These tables provide the median DEFFs for the 8 large States and the median of the 43 State medians for the remaining States. Results for the smaller States are provided for reference only. Although these DEFFs were of the same order as those for the larger States (because the sample design was the same for all States), the above approximate formula is not recommended for use with smaller States because of the instability of the prevalence estimates. The small area estimation (SAE) methodology was used, as in the case of NSDUH reports since 1999, to improve the precision of the State estimates. To get an idea of the magnitude of the 2012 drug-specific DEFFs used in computing the median DEFF over the drugs, Table 6.4 lists the 56 individual DEFFs for each of the age groups and the national total.

[18] The demographic variables for county type and race underwent minor revisions in 2003 and remained revised for 2012. The Office of Management and Budget, whose definitions are used to determine county type, changed the way "metropolitan" was defined. Counties may have moved between levels of county type specifically because of this change. For race, there is no longer a response in the questionnaire that allows respondents to choose a "main race." Instead "main race" was imputed for multiple race respondents.

Table 6.1 Median Design Effects of Lifetime Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
Total	1.69	2.10	1.92	3.69	3.38	3.70
Gender						
Male	1.75	1.87	1.85	N/A	N/A	3.69
Female	1.74	2.10	1.75	N/A	N/A	3.38
Age (Years)						
12–17	N/A	N/A	N/A	1.75	1.74	1.69
18–25	N/A	N/A	N/A	1.87	2.10	2.10
26+	N/A	N/A	N/A	1.85	1.75	1.92
Race/Ethnicity						
White, Not Hispanic or Latino	1.57	1.99	1.87	3.33	3.35	3.54
Black or African American, Not Hispanic or Latino	1.49	1.76	2.09	4.11	3.18	4.00
Hispanic or Latino	2.07	2.13	2.15	4.74	3.63	4.20
Other or Multiple, Not Hispanic or Latino	1.80	2.59	1.98	4.56	2.80	3.99
Population Density						
Large Metropolitan	1.58	2.03	1.75	3.41	3.11	3.34
Small Metropolitan	1.63	2.06	2.00	3.87	3.59	3.99
Nonmetropolitan	1.84	1.97	1.81	3.05	2.92	3.45
Census Division						
New England	2.23	3.06	1.86	4.34	3.20	3.67
Middle Atlantic	1.43	1.84	1.66	3.27	2.97	3.27
East North Central	1.35	1.76	1.51	2.92	2.46	2.98
West North Central	1.33	2.40	2.14	3.74	3.71	4.49
South Atlantic	1.67	1.97	1.73	3.48	3.28	3.58
East South Central	1.23	1.36	1.56	2.88	2.45	3.03
West South Central	1.57	1.35	1.41	2.71	2.37	2.69
Mountain	2.04	2.37	2.38	4.48	3.46	4.61
Pacific	1.51	2.03	1.83	3.56	3.22	3.48
County Type[1]						
Large Metropolitan	1.61	2.05	1.80	3.52	3.18	3.42
Small Metropolitan I	1.54	1.86	1.86	3.71	3.70	3.71
Small Metropolitan II	1.86	2.46	2.04	4.40	3.37	4.09
Nonmetropolitan I	1.50	1.93	1.94	3.41	3.03	3.63
Nonmetropolitan II	1.65	1.94	1.74	3.46	2.35	3.42
Nonmetropolitan III	2.34	1.95	1.95	4.10	2.48	3.60
Adult Education[2]						
Less Than High School	N/A	1.93	1.69	2.46	2.25	2.58
High School Graduate	N/A	1.91	1.83	2.97	2.56	2.88
Some College	N/A	2.02	1.85	2.77	2.88	2.90
College Graduate	N/A	1.97	2.05	2.39	2.53	2.60
Current Employment[3]						
Full-Time	N/A	1.84	1.96	2.57	2.57	2.61
Part-Time	N/A	2.12	1.88	3.42	3.22	3.32
Unemployed	N/A	1.88	1.77	3.35	2.89	3.07
Other[4]	N/A	1.97	1.68	2.39	2.19	2.37

(continued)

Table 6.1 Median Design Effects of Lifetime Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012 (continued)

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
State						
California	1.18	1.55	1.42	2.77	2.49	2.78
Florida	1.35	1.49	1.53	3.09	2.92	3.41
Illinois	1.10	1.30	1.46	2.54	2.14	2.61
Michigan	1.25	1.47	1.15	2.52	1.97	2.29
New York	1.23	1.33	1.49	2.91	2.56	2.96
Ohio	1.10	1.39	1.18	2.49	2.05	2.43
Pennsylvania	1.03	1.37	1.61	2.97	2.71	3.29
Texas	1.38	1.27	1.35	2.60	2.25	2.60
Other[5]	1.04	1.36	1.39	2.68	2.18	2.66

N/A = not applicable.

Note: These design effects apply to the following drugs: any illicit drug use, marijuana/hashish, cocaine, crack, inhalants, hallucinogens, LSD, PCP, heroin, nonmedical use of any psychotherapeutics, nonmedical use of stimulants, nonmedical use of sedatives, nonmedical use of tranquilizers, nonmedical use of pain relievers, and any illicit drug except marijuana.

[1]Data on County Type are defined as follows:

Large Metropolitan: counties in metro areas with a population ≥ 1 million.

Small Metropolitan I: counties in metro areas with a population between 250,000 and 1 million.

Small Metropolitan II: counties in metro areas with a population < 250,000.

Nonmetropolitan I: urban populations not part of metro areas ≥ 20,000.

Nonmetropolitan II: urban populations not part of metro areas between 2,500 and 19,999.

Nonmetropolitan III: completely rural.

[2]Data on adult education are not applicable for persons aged 12 to 17.

[3]Data on current employment are not applicable for persons aged 12 to 17.

[4]Retired, disabled, homemaker, student, or "other."

[5]Median of the median design effects for the 43 States.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.2 Median Design Effects of Past Year and Past Month Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
Total	1.68	2.07	1.74	2.73	2.03	2.58
Gender						
Male	1.74	1.96	1.65	N/A	N/A	2.73
Female	1.77	1.96	1.50	N/A	N/A	2.03
Age (Years)						
12–17	N/A	N/A	N/A	1.74	1.77	1.68
18–25	N/A	N/A	N/A	1.96	1.96	2.07
26+	N/A	N/A	N/A	1.65	1.50	1.74
Race/Ethnicity						
White, Not Hispanic or Latino	1.65	2.07	1.63	2.43	1.81	2.48
Black or African American, Not Hispanic or Latino	1.52	1.75	1.62	3.32	1.37	2.52
Hispanic or Latino	1.97	2.04	1.61	2.46	2.34	2.49
Other or Multiple, Not Hispanic or Latino	1.67	2.48	1.58	2.37	1.56	2.65
Population Density						
Large Metropolitan	1.65	1.95	1.50	2.40	1.83	2.37
Small Metropolitan	1.56	1.83	1.84	2.90	2.05	2.93
Nonmetropolitan	1.59	1.75	1.21	1.53	1.30	1.63
Census Division						
New England	2.07	2.40	1.40	2.46	1.62	1.91
Middle Atlantic	1.40	1.88	1.36	2.38	1.73	2.28
East North Central	1.24	1.85	1.19	1.75	1.65	1.85
West North Central	1.37	1.85	1.49	2.00	1.42	2.21
South Atlantic	1.64	1.84	1.58	2.76	1.86	2.55
East South Central	1.29	1.25	1.11	1.87	1.00	1.64
West South Central	1.49	1.21	1.10	1.73	1.39	1.84
Mountain	1.64	1.88	1.78	2.29	2.55	2.63
Pacific	1.50	1.99	1.58	2.30	1.92	2.28
County Type[1]						
Large Metropolitan	1.67	1.98	1.56	2.49	1.91	2.46
Small Metropolitan I	1.40	1.81	1.43	2.29	1.56	2.49
Small Metropolitan II	1.75	1.86	1.40	2.54	1.76	2.17
Nonmetropolitan I	1.52	1.68	1.35	2.32	1.16	2.15
Nonmetropolitan II	1.35	1.60	1.17	1.74	1.12	1.69
Nonmetropolitan III	2.30	1.53	1.34	1.79	1.01	1.87
Adult Education[2]						
Less Than High School	N/A	1.75	1.49	1.70	1.62	1.84
High School Graduate	N/A	1.86	1.62	2.19	1.21	1.99
Some College	N/A	2.16	1.52	2.09	1.67	2.10
College Graduate	N/A	1.77	1.58	1.80	1.93	1.84
Current Employment[3]						
Full-Time	N/A	1.91	1.67	1.97	1.77	2.02
Part-Time	N/A	2.19	1.48	2.28	1.83	2.25
Unemployed	N/A	1.83	1.54	2.65	1.23	2.43
Other[4]	N/A	1.97	1.44	1.87	1.39	1.76

(continued)

Table 6.2 Median Design Effects of Past Year and Past Month Illicit Drug Use, by Age Group, Gender, and Demographic Characteristic: 2012 (continued)

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
State						
California	1.22	1.53	1.22	1.75	1.43	1.80
Florida	1.34	1.46	1.15	1.56	1.54	1.73
Illinois	1.07	1.44	1.28	1.78	1.29	1.79
Michigan	1.25	1.34	1.00	1.65	1.29	1.54
New York	1.21	1.20	1.31	1.75	1.33	1.78
Ohio	1.00	1.22	1.00	1.53	1.55	1.57
Pennsylvania	1.14	1.50	1.27	2.03	1.50	1.83
Texas	1.31	1.10	1.07	1.56	1.34	1.66
Other[5]	1.08	1.25	1.11	1.36	1.05	1.49

N/A = not applicable.

Note: These design effects apply to the following drugs: any illicit drug use, marijuana/hashish, cocaine, crack, inhalants, hallucinogens, LSD, PCP, heroin, nonmedical use of any psychotherapeutics, nonmedical use of stimulants, nonmedical use of sedatives, nonmedical use of tranquilizers, nonmedical use of pain relievers, and any illicit drug except marijuana.

[1]Data on County Type are defined as follows:

Large Metropolitan: counties in metro areas with a population ≥ 1 million.

Small Metropolitan I: counties in metro areas with a population between 250,000 and 1 million.

Small Metropolitan II: counties in metro areas with a population < 250,000.

Nonmetropolitan I: urban populations not part of metro areas ≥ 20,000.

Nonmetropolitan II: urban populations not part of metro areas between 2,500 and 19,999.

Nonmetropolitan III: completely rural.

[2]Data on adult education are not applicable for persons aged 12 to 17.

[3]Data on current employment are not applicable for persons aged 12 to 17.

[4]Retired, disabled, homemaker, student, or "other."

[5]Median of the median design effects for the 43 States.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.3 Median Design Effects of Lifetime, Past Year, and Past Month Licit Drug Use Estimates, by Age Group, Gender, and Demographic Characteristic: 2012

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
Total	1.78	2.28	2.11	3.45	3.49	3.63
Gender						
Male	1.83	2.09	1.91	N/A	N/A	3.45
Female	1.63	2.10	2.01	N/A	N/A	3.49
Age (Years)						
12–17	N/A	N/A	N/A	1.83	1.63	1.78
18–25	N/A	N/A	N/A	2.09	2.10	2.28
26+	N/A	N/A	N/A	1.91	2.01	2.11
Race/Ethnicity						
White, Not Hispanic or Latino	1.67	2.06	2.00	3.06	2.65	3.16
Black or African American, Not Hispanic or Latino	1.60	1.68	2.09	3.67	3.67	3.95
Hispanic or Latino	1.99	2.30	2.09	4.03	3.43	4.06
Other or Multiple, Not Hispanic or Latino	1.80	2.26	2.89	5.20	3.79	5.40
Population Density						
Large Metropolitan	1.68	2.01	1.79	3.14	3.00	3.47
Small Metropolitan	1.84	2.21	2.34	3.50	3.82	3.87
Nonmetropolitan	2.38	2.16	2.40	3.88	3.83	4.28
Census Division						
New England	1.77	3.13	2.54	3.91	3.71	4.09
Middle Atlantic	1.48	1.81	1.66	3.02	2.69	3.06
East North Central	1.31	1.62	1.74	2.57	2.95	3.25
West North Central	1.70	2.45	2.41	3.66	3.43	4.40
South Atlantic	2.07	2.26	1.88	3.27	3.78	3.50
East South Central	1.51	1.56	1.34	2.65	2.56	2.58
West South Central	1.53	1.73	1.61	2.47	2.66	2.99
Mountain	2.10	2.56	2.47	4.13	3.49	4.46
Pacific	1.54	2.08	2.10	3.43	2.70	3.61
County Type[1]						
Large Metropolitan	1.67	2.04	1.81	3.15	2.99	3.47
Small Metropolitan I	1.58	1.96	2.12	3.21	3.57	3.70
Small Metropolitan II	1.99	2.42	2.07	3.52	3.60	4.20
Nonmetropolitan I	1.56	2.14	2.29	3.64	3.59	3.92
Nonmetropolitan II	2.51	2.60	2.14	3.78	3.29	3.96
Nonmetropolitan III	2.44	1.97	2.27	4.16	5.28	4.40
Adult Education[2]						
Less Than High School	N/A	1.93	2.07	3.03	3.06	3.15
High School Graduate	N/A	2.04	1.83	2.65	2.85	2.84
Some College	N/A	2.11	1.81	2.86	2.47	2.75
College Graduate	N/A	1.84	1.95	2.24	2.15	2.47
Current Employment[3]						
Full-Time	N/A	2.00	1.96	2.57	2.37	2.64
Part-Time	N/A	2.03	1.74	3.03	2.63	2.79
Unemployed	N/A	1.80	1.82	3.12	3.03	2.97
Other[4]	N/A	2.42	1.88	2.54	2.87	2.72

(continued)

Table 6.3 Median Design Effects of Lifetime, Past Year, and Past Month Licit Drug Use Estimates, by Age Group, Gender, and Demographic Characteristic: 2012 (continued)

Demographic Characteristic	Age Group			Gender		Total
	12–17	18–25	26+	Male	Female	
State						
California	1.19	1.63	1.55	2.79	2.36	2.95
Florida	1.42	1.34	1.43	2.51	3.05	2.87
Illinois	1.29	1.56	1.26	2.31	2.03	2.54
Michigan	1.00	1.44	1.21	2.13	2.01	2.16
New York	1.38	1.61	1.63	2.95	3.07	3.14
Ohio	1.19	1.36	1.40	2.41	2.06	2.51
Pennsylvania	1.21	1.28	1.50	2.48	2.23	2.53
Texas	1.40	1.61	1.47	2.37	2.24	2.78
Other[5]	1.26	1.50	1.48	2.36	2.24	2.64

N/A = not applicable.

Note: These design effects apply to the following substances: cigarettes, alcohol, binge drinking, heavy drinking, and smokeless tobacco.

[1]Data on County Type are defined as follows:

Large Metropolitan: counties in metro areas with a population ≥ 1 million.

Small Metropolitan I: counties in metro areas with a population between 250,000 and 1 million.

Small Metropolitan II: counties in metro areas with a population < 250,000.

Nonmetropolitan I: urban populations not part of metro areas ≥ 20,000.

Nonmetropolitan II: urban populations not part of metro areas between 2,500 and 19,999.

Nonmetropolitan III: completely rural.

[2]Data on adult education are not applicable for persons aged 12 to 17.

[3]Data on current employment are not applicable for persons aged 12 to 17.

[4]Retired, disabled, homemaker, student, or "other."

[5]Median of the median design effects for the 43 States.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

Table 6.4 Design Effects, by Age Group, for the Outcomes Used in the Medians in Tables 6.1, 6.2, and 6.3: 2012

Outcome	Age Group			Total
	12–17	18–25	26+	
Illicit Drugs, Lifetime Recency				
Any Illicit Drug	1.68	2.36	1.98	3.70
Marijuana	1.71	2.38	2.20	4.16
Cocaine	1.87	2.22	1.96	4.02
Crack	2.49	1.63	1.72	3.62
Inhalants	1.57	1.74	1.69	3.24
Hallucinogens	1.69	2.23	1.82	3.69
LSD	1.69	2.55	1.76	3.83
PCP	1.50	1.74	1.87	3.99
Heroin	2.16	1.84	1.79	3.51
Nonmedical Use of Psychotherapeutics	1.64	2.11	1.92	3.58
Nonmedical Use of Stimulants	1.57	2.12	2.15	4.19
Nonmedical Use of Sedatives	1.44	2.06	2.30	4.84
Nonmedical Use of Tranquilizers	1.92	2.01	2.01	3.87
Nonmedical Use of Pain Relievers	1.60	2.06	1.71	3.20
Any Illicit Drug except Marijuana	1.72	2.10	1.92	3.67
Illicit Drugs, Past Year Recency				
Any Illicit Drug	1.69	2.46	1.84	3.17
Marijuana	1.74	2.55	1.75	2.93
Cocaine	1.89	2.36	1.63	2.71
Crack	2.93	1.43	1.84	3.56
Inhalants	1.65	1.99	2.34	2.27
Hallucinogens	1.66	2.48	1.48	1.91
LSD	1.58	2.56	1.75	1.79
PCP	2.15	1.50	1.53	1.73
Heroin	2.33	1.97	1.67	2.42
Nonmedical Use of Psychotherapeutics	1.64	1.97	1.74	2.82
Nonmedical Use of Stimulants	1.49	2.60	1.46	2.06
Nonmedical Use of Sedatives	1.43	1.73	1.18	1.92
Nonmedical Use of Tranquilizers	2.01	2.01	1.67	2.81
Nonmedical Use of Pain Relievers	1.61	1.95	1.75	2.79
Any Illicit Drug except Marijuana	1.77	2.07	1.74	2.84

(continued)

Table 6.4 Design Effects, by Age Group, for the Outcomes Used in the Medians in Tables 6.1, 6.2, and 6.3: 2012 (continued)

Outcome	Age Group			Total
	12–17	18–25	26+	
Illicit Drugs, Past Month Recency				
Any Illicit Drug	1.69	2.13	1.81	2.94
Marijuana	1.64	2.20	1.78	2.79
Cocaine	1.40	1.77	1.82	3.25
Crack	*	1.31	2.53	5.26
Inhalants	1.74	2.56	2.02	2.24
Hallucinogens	1.72	2.32	1.83	1.90
LSD	1.15	1.70	1.94	1.39
PCP	1.61	1.07	1.00	1.32
Heroin	2.32	1.23	*	2.01
Nonmedical Use of Psychotherapeutics	1.78	2.08	1.71	2.79
Nonmedical Use of Stimulants	1.55	1.65	1.67	2.33
Nonmedical Use of Sedatives	1.41	2.08	1.34	2.39
Nonmedical Use of Tranquilizers	2.42	2.21	1.92	3.26
Nonmedical Use of Pain Relievers	1.68	2.19	1.57	2.46
Any Illicit Drug except Marijuana	1.90	2.07	1.74	2.92
Licit Drugs, Lifetime Recency				
Alcohol	2.02	2.29	2.25	3.16
Cigarettes	1.78	2.28	2.21	3.92
Smokeless Tobacco	1.72	2.05	1.83	3.45
Licit Drugs, Past Year Recency				
Alcohol	2.05	2.33	2.36	4.15
Cigarettes	1.78	2.15	2.13	4.15
Smokeless Tobacco	1.68	1.75	1.45	2.42
Licit Drugs, Past Month Recency				
Alcohol	1.95	2.28	2.22	4.18
Cigarettes	1.78	2.26	2.11	4.12
Smokeless Tobacco	1.67	1.75	1.49	2.68
Binge Drinking	1.68	2.42	1.89	3.63
Heavy Drinking	1.76	2.47	1.77	3.32

*Low precision; no estimate reported.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2012.

7. Conclusions

As stated in Chapter 1, it is important to have a good understanding of the resulting standard errors (SEs) and design effects (DEFFs) corresponding to a set of key outcome variables and other variables as part of any survey data analysis. One reason for this is to evaluate how well the sample was designed in light of the target and observed precisions, and the DEFFs. The 2012 National Survey on Drug Use and Health (NSDUH) met its precision goals for all four target domains defined by three age groups (12 to 17, 18 to 25, 26 or older, and total [i.e., 12 or older]).

Another important reason for the examination of SEs and DEFFs is to obtain quick estimates of SEs for any user-specified outcome variable through some form of modeling. Although SEs of several prevalence estimates are available from published analysis reports on the survey, SEs of other estimates of interest by a user may not be available in the published tables. If the user has access to the primary data source (i.e., micro data), the SE can be computed using commercially available software, such as SUDAAN®. However, a user often has access to only a secondary data source without information on SEs included. For this case, it is useful to have a provision for computing quick and approximate SEs. If the secondary data source contains information about median DEFFs (over a set of drug use variables) for selected demographic domains, such as age and race/ethnicity, a rough approximate SE can be obtained easily using the formula (Equation 4) for variance as a function of DEFF, domain sample size, and the prevalence estimate. The formula is

$$var\left(p_d\right)_{appx} = DEFF_{d,MED} * \left[p_d\left(1-p_d\right)/n_d\right].$$

Note that the use of a known median DEFF in place of a variable-specific unknown DEFF provides a simple type of modeling. One also could use mean DEFF instead of median DEFF. This report contains tables showing median and mean DEFFs for a number of domains. The differences are generally small. The above way of modeling SEs, via median DEFF, is not applicable if the available median DEFF does not correspond to the domain of interest.

In summary, the user may obtain SE estimates for the 2012 NSDUH for drug recency outcomes from the following sources in recommended order:

1. commercially available variance estimation software packages, such as SUDAAN;
2. published SEs from reports using data from the 2012 NSDUH (to be available at http://www.samhsa.gov/data/NSDUH.aspx or upon request from the Center for Behavioral Health Statistics and Quality at the Substance Abuse and Mental Health Services Administration);
3. median domain DEFFs appearing in Tables 6.1, 6.2, and 6.3 and application of Equation 4 in Chapter 6 for drug recency of use; and

4. online data analysis system in the Substance Abuse and Mental Health Data Archive (SAMHDA), available at http://www.icpsr.umich.edu/icpsrweb/SAMHDA/, provided that the stratum and primary sampling unit variables are specified.

References

Center for Behavioral Health Statistics and Quality. (in press). *Results from the 2012 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 13-xxxx, NSDUH Series H-x). Rockville, MD: Substance Abuse and Mental Health Services Administration. To be available at http://www.samhsa.gov/data/NSDUH.aspx

Chromy, J. R. (1979). Sequential sample selection methods. In *Proceedings of the 1979 American Statistical Association, American Statistical Association, Survey Research Methods Section, Washington, DC* (pp. 401-406). Washington, DC: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Chromy, J. R., & Penne, M. (2002). Pair sampling in household surveys. In *Proceedings of the 2002 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, New York, NY* (pp. 552-554). Alexandria, VA: American Statistical Association. [Available as a PDF at http://www.amstat.org/sections/SRMS/proceedings/]

Kish, L. (1965). *Survey sampling* (pp. 313-315). New York: John Wiley & Sons.

Morton, K. B., Martin, P. C., Shook-Sa, B.E., Chromy, J. R, & Hirsch, E. L. (2013). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, Phase II, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International. To be available at http://www.samhsa.gov/data/Methodological_Reports.aspx

Office of Management and Budget. (2003, June 6). *OMB Bulletin No. 03-04: Revised definitions of metropolitan statistical areas, new definitions of micropolitan statistical areas and combined statistical areas, and guidance on uses of the statistical definitions of these areas*. Retrieved January 29, 2013, from http://www.whitehouse.gov/omb/bulletins_b03-04

RTI International. (2008). *SUDAAN®, Release 10.0* [computer software]. Research Triangle Park, NC: Author.

U.S. Census Bureau, Redistricting Data Office. (2009). Appendix A: Geographic terms and concepts. In *2008 Redistricting data prototype (Public Law 94-171) summary file*. Washington, DC: U.S. Census Bureau. [Available as a PDF at http://www.census.gov/rdo/; see http://www.census.gov/rdo/pdf/pl94-171.pdf]

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

METHODOLOGICAL RESOURCE BOOK SECTION 16a: MENTAL HEALTH SURVEILLANCE STUDY OPERATIONS REPORT (2008-2012)

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland 20857

February 2014

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH: MENTAL HEALTH SURVEILLANCE STUDY OPERATIONS REPORT (2008-2012)

Prepared for the 2012 Methodological Resource Book (Section 16a)

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.112.001
Deliverable No. 41

RTI Authors:

Rebecca Granger
Ryan Blazei
Lee Ellen Coffey
Ryan Gordon
Rhonda Karg
Kate Panzer
Bonnie Shook-Sa

SAMHSA Authors:

Peggy Barker
Jonaki Bose
Joe Gfroerer
Sarra Hedden

RTI Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

February 2014

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2014). *2012 National Survey on Drug Use and Health: Methodological Resource Book (Section 16a, Mental Health Surveillance Study Operations Report [2008-2012])*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This report would not be possible without the guidance and input of staff from the Substance Abuse and Mental Health Services Administration (SAMHSA). Special thanks are also due to several RTI International (a trade name of Research Triangle Institute) staff members. Significant contributors include Dan Liao and Susan Myers. Claudia Clark and Margaret Smith copyedited the report, and Valerie Garner formatted the report in preparation for publication. Teresa Bass provided web production.

DISCLAIMER

SAMHSA provides links to other Internet sites as a service to its users and is not responsible for the availability or content of these external sites. SAMHSA, its employees, and contractors do not endorse, warrant, or guarantee the products, services, or information described or offered at these other Internet sites. Any reference to a commercial product, process, or service is not an endorsement or recommendation by SAMHSA, its employees, or contractors. For documents available from this server, the U.S. Government does not warrant or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed.

Table of Contents

Chapter **Page**

1. Introduction 1

2. Management Structure and Staffing 5
   - 2.1 Project Management 6
   - 2.2 Instrument Assessment and Development 6
   - 2.3 Sample Design and Selection 7
   - 2.4 Training and Field Preparation 7
   - 2.5 Data Collection 8
   - 2.6 Data Management 9
   - 2.7 Data Analysis and Reporting 9
   - 2.8 Data Files and Documentation 10

3. Data Collection Staffing 11
   - 3.1 Data Collection Managers 11
   - 3.2 Clinical Supervisors 12
   - 3.3 Clinical Interviewers 12
   - 3.4 Problems Encountered 14
     - 3.4.1 CI Turnover 15

4. Data Collection Materials 17
   - 4.1 MHSS Recruitment Materials 17
     - 4.1.1 CAI Recruitment Scripts 17
     - 4.1.2 Respondent Materials 17
   - 4.2 MHSS Interview Materials 17
     - 4.2.1 Introduction and Informed Consent 18
     - 4.2.2 SCID Content 18
     - 4.2.3 GAF Score 21
     - 4.2.4 Short Blessed Scale 21
     - 4.2.5 Distressed Respondent Protocol 22
     - 4.2.6 Interviewer Debriefing 23
   - 4.3 CI Equipment 23
     - 4.3.1 Laptop 23
     - 4.3.2 Audio Recording 24
     - 4.3.3 Case Management System 24
   - 4.4 MHSS Handbooks 24
     - 4.4.1 MHSS CI Handbook 25
     - 4.4.2 MHSS FI Handbook 26
   - 4.5 OMB and IRB Packages 26
     - 4.5.1 OMB Package 26
     - 4.5.2 IRB Package 26

Table of Contents (continued)

Chapter **Page**

5. Staff Training 29
   5.1 MHSS Recruitment FI Training 29
      5.1.1 Veteran FI Training 29
      5.1.2 NTP FI Training 29
      5.1.3 FI Refresher Training 30
   5.2 NTP CI Training Sessions 30
      5.2.1 Training Staff 30
      5.2.2 Content of NTP CI Training Sessions 31
      5.2.3 Certifications 32
   5.3 Veteran CI Training Sessions 33
      5.3.1 Staffing 33
      5.3.2 Content of Veteran CI Training Sessions 33
   5.4 IR Exercises 35
   5.5 CS Training 37

6. Data Collection 39
   6.1 Schedule 39
   6.2 SCID Administration 39
      6.2.1 SCID Length 41
   6.3 SCID Data Processing 42
      6.3.1 Clinical Editing 43
      6.3.2 Technical Editing 43
      6.3.3 Keying 44
      6.3.4 Storage and Destruction 44
   6.4 Data Collection Management and Reporting 45
      6.4.1 MHSS Sample Monitoring Report 46
      6.4.2 Clinical Interview Current Status Report 46
   6.5 Problems Encountered 46

7. Data Collection Results 49
   7.1 Response Rates 49
   7.2 Number of Contacts 53
   7.3 Distressed Respondents 59
   7.4 Short Blessed Scale 62
   7.5 Problems Encountered 63

8. Data Collection Quality Control 65
   8.1 Clinical Supervision 65
   8.2 Invalid Cases 67
      8.2.1 Invalid Clinical Interviews 67
      8.2.2 Invalid Main Study Interviews 69

References 71

Table of Contents (continued)

Appendices

A CAI Recruitment Scripts

B Recruitment Materials

C Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)

D Mental Health Surveillance Study Clinical Interviewer Handbook

E Mental Health Surveillance Study Field Interviewer Handbook

F New Clinical Interviewer Training Agenda

G SCID Editing Form

H SCID Technical Editing

I Clinical Interview Current Status Report

List of Tables

Table		Page
3.1	Number of Clinical Interviewers Staffed for the 2008-2012 MHSS	12
3.2	MHSS Clinical Interviewer Training and Certification Results	15
4.1	Diagnostic Modules Contained in the MHSS SCID	19
4.2	Distressed Respondent Protocol Scenarios	22
5.1	Interrater Reliability Exercise Performance	37
6.1	NSDUH and MHSS Data Collection End Date, by Year	39
6.2	2008-2012 NSDUH MHSS SCID Length	42
6.3	MHSS Sample Data: 2012 Quarter 4	46
7.1	2008-2012 MHSS Response Rate Summary	50
7.2	2008 MHSS, Quarters 1 through 4 Summary	50
7.3	2009 MHSS, Quarters 1 through 4 Summary	51
7.4	2010 MHSS, Quarters 1 through 4 Summary	51
7.5	2011 MHSS, Quarters 1 through 4 Summary	52
7.6	2012 MHSS, Quarters 1 through 4 Summary	52
7.7	Number of Call Attempts Made To Contact MHSS Respondents: 2008-2012	54
7.8	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2008-2012	54
7.9	Number of Call Attempts Made To Contact MHSS Respondents: 2008	55
7.10	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2008	55
7.11	Number of Call Attempts Made To Contact MHSS Respondents: 2009	56
7.12	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2009	56
7.13	Number of Call Attempts Made To Contact MHSS Respondents: 2010	57
7.14	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2010	57
7.15	Number of Call Attempts Made To Contact MHSS Respondents: 2011	58
7.16	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2011	58
7.17	Number of Call Attempts Made To Contact MHSS Respondents: 2012	59
7.18	Number of Call Attempts Made To Contact MHSS Nonrespondents: 2012	59
7.19	Distressed Respondent Protocol Scenarios: 2008-2012 MHSS	60
7.20	Number of Distressed Respondent Protocol Classifications: 2008-2012 MHSS	60
7.21	Number of Institutional Review Board Violations, by Distressed Respondent Protocol Scenarios: 2008-2012 MHSS	61

List of Tables (continued)

Table **Page**

7.22 Number and Percentage of Cases with Institutional Review Board Violations, by Distressed Respondent Protocol Scenarios: 2008-2012 MHSS 61

7.23 Demographic Characteristics of Respondents Who Exceeded the SBS Cutoff Score: 2008-2012 MHSS 62

7.24 MHSS Interviews Associated with Main Study Interviews Completed, by Pennsylvania Field Interviewer: 2008-2011 63

8.1 Types of Discarded Cases, by Year: 2008-2012 MHSS.....69

List of Exhibits

Exhibit **Page**

2.1 MHSS Organizational Structure 5

1. Introduction

The overarching goal of the Mental Health Surveillance Study (MHSS) of the National Survey on Drug Use and Health (NSDUH) is to provide accurate estimates of the prevalence of serious mental illness (SMI) among adults aged 18 or older at the national and State levels. Public Law No. 102-321, the Alcohol, Drug Abuse, and Mental Health Administration Reorganization Act of 1992, established a block grant for States to fund community mental health services for adults with SMI. The law required States to include prevalence estimates in their annual applications for block grant funds. This legislation also required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop an operational definition of SMI and to produce national and State estimates. The MHSS clinical follow-up study was conducted as part of NSDUH to develop a model to generate estimates of SMI.

On May 20, 1993, SAMHSA's Center for Mental Health Services (CMHS) published its definition of SMI in the *Federal Register*:

> Pursuant to Section 1912(c) of the Public Health Services Act, as amended by Public Law 102-321, "adults with serious mental illness" are defined as the following:
>
> - Persons aged 18 and over, who currently or at any time during the past year, have had diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM-III-R [sic] that has resulted in functional impairment, which substantially interferes with or limits one or more major life activities.
> - These disorders include any mental disorders (including those of biological etiology) listed in DSM-III-R or their ICD-9-CM equivalent (and subsequent revisions), with the exception of DSM-III-R "V" codes, substance use disorders, and developmental disorders, which are excluded unless they co-occur with other diagnosable serious mental illness.
> - All of these disorders have episodic, recurrent, or persistent features; however, they vary in terms of severity or disabling effects. Functional impairment is defined as difficulties that substantially interfere with or limit role functioning in one or more major life activities, including basic daily living skills (e.g., eating, bathing, dressing); instrumental living skills (e.g., maintaining a household, managing money, getting around the community, taking prescribed medication); and functioning in social, family, and vocational/educational contexts.
> - Adults who would have met functional impairment criteria during the referenced year without benefit of treatment or other support services are considered to have serious mental illnesses.

In December 2006, a technical advisory group meeting of expert consultants was convened by the Office of Applied Studies (now the Center for Behavioral Health Statistics and Quality [CBHSQ]) and CMHS to solicit recommendations for mental health surveillance data collection strategies among the U.S. population. The panel recommended that NSDUH be used to produce estimates of SMI among adults by including short scales in NSDUH's main interview

that are strong predictors of SMI and that a "gold standard" clinical psychiatric interview be administered to a subset of respondents to provide the data for estimating a statistical model that predicts SMI. In response, SAMHSA's CBHSQ initiated the MHSS as part of NSDUH to develop and implement a method to estimate SMI. At the time, NSDUH contained a six-item scale (Kessler-6, or K6) with five response options in each item that captured information on psychological distress in the past 12 months (Kessler et al., 2003). However, the K6 scale is not a diagnostic instrument and does not capture information on functional impairment, which is needed to determine whether a respondent can be categorized as having SMI under SAMHSA's definition. In consultation with the technical advisory group, two candidate impairment scales were selected by SAMHSA to be added to the 2008 NSDUH. They were an abridged version of the World Health Organization Disability Assessment Schedule (WHODAS; Rehm et al., 1999) and the Sheehan Disability Scale (SDS; Leon, Olfson, Portera, Farber, & Sheehan, 1997). An initial step was to modify these scales for use in a general population survey, including changes to question wording and length, which resulted in an abbreviated eight-item version of the WHODAS (Novak, Colpe, Barker, & Gfroerer, 2010).

A feasibility study was conducted in June 2007 to test the procedures and instruments proposed for the MHSS, as well as to assess the training needs for both the field interviewers and the clinical interviewers.[1] The MHSS was conducted first in 2008. A split-sample design was used in the 2008 NSDUH, for which all adult respondents to the main NSDUH interview received the K6, but a random half of the sample received the WHODAS and the other half received the SDS. In addition, a subsample of approximately 1,500 adult NSDUH participants completed a follow-up clinical interview to provide data for developing models to estimate mental illness using the NSDUH full-sample interview data. The randomization of the impairment scales was maintained within this clinical interview subsample, so that about half of the MHSS sample participants were administered the WHODAS and half were administered the SDS (i.e., there were approximately 750 completed interviews from each half sample). Each participant in the 2008 MHSS clinical study was administered the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002), which was adapted for this study. Trained clinical interviewers administered the paper-and-pencil SCID to respondents over the telephone approximately 2 to 4 weeks after the NSDUH interview. Functional impairment ratings were assigned by clinical interviewers using the Global Assessment of Functioning (GAF) scale.[2] The model estimation analyses used "gold standard" measures (i.e., the SCID/GAF combination as the indicator of SMI) in evaluating which combination of the K6 and impairment scale worked best in the statistical model used to predict SMI status. Based on an analysis of the 2008 MHSS data, it was determined that the WHODAS would be administered as the sole impairment scale in subsequent NSDUHs (starting in 2009) and that it would be used in combination with the K6

[1] More details on the MHSS feasibility study can be found in Section 3.3 of the 2012 MHSS design and estimation report (Liao et al., in press).

[2] The GAF score is based on a numeric scale (0 through 100) that is used to subjectively rate the social, occupational, and psychological functioning of adults and is presented and described in the DSM-IV-TR (American Psychiatric Publishing, 2000, p. 32; Endicott, Spitzer, Fleiss, & Cohen, 1976). Lower scores represent higher levels of functional impairment. Descriptions of impairment are provided at 10-point intervals such as 1 to 10, 11 to 20, and so on up to 91 to 100. For example, a GAF score between 51 and 60 is described as having moderate symptoms of impairment, while a score higher than 60 represents several categories of impairment ranging from none to slight, and a score lower than 51 represents several categories ranging from serious to extreme.

scale to predict SMI. For more details, refer to the 2008 MHSS analysis report by Aldworth et al. (2009).

From 2009 through 2012, the MHSS was conducted similarly to the 2008 MHSS, with two major differences: (1) only the WHODAS impairment scale was administered in the NSDUH, and (2) the sample size was approximately 500 in 2009 and 2010 and approximately 1,500 in 2011 and 2012. Based on the 2008 MHSS clinical data, a model was developed and implemented in the 2008-2011 NSDUHs to produce a predicted probability of having SMI for each clinical interview respondent. Although this model was optimized to predict SMI, the SMI predicted probabilities were also used to predict mild (or low) mental illness, moderate mental illness, and any mental illness (AMI) using different cut points (for more details about this model, see Liao et al., 2012). After data collection ended for the 2012 MHSS, a revised and improved model was used to predict mental illness for the 2012 NSDUH using the combined clinical data from 2008 to 2012 (subsequently referred to as the "2012 model"). The 2008 model substantially overestimated SMI and AMI among young adults relative to the clinical interview data because of the small number of respondents in the 2008 clinical subsample. In addition, improvements were needed in the weighting procedures for the MHSS clinical data to better account for nonresponse and undercoverage (i.e., because persons with mental illness appeared to be more likely to participate in the clinical follow-up and because only NSDUH respondents who answered their surveys in English were eligible for the follow-up). To reduce bias and improve prediction, additional mental health–related variables and an age variable were added in the 2012 model. To reduce coverage and nonresponse error, alternatives for the weights were applied to the clinical sample data for the model development. To provide consistent data for trend assessment, the mental illness estimates were recomputed for 2008-2011 using the new 2012 model. The 2012 MHSS design and estimation report contains the details on the estimation of mental illness (Liao et al., in press).

This report describes data collection operations for the 2008-2012 MHSS. Specifically, the following topics are described: Management Structure and Staffing, Data Collection Staffing, Data Collection Materials, Staff Training, Data Collection, Data Collection Results, and Data Collection Quality Control. Unless changes are specified in this report, the operations, protocols, and materials remained generally the same for the MHSS each year.

2. Management Structure and Staffing

This chapter describes the management organizational structure and staffing for the Mental Health Surveillance Study (MHSS) task on the National Survey on Drug Use and Health (NSDUH) contract. NSDUH is sponsored by the Center for Behavioral Health Statistics and Quality (formerly the Office of Applied Studies) within the Substance Abuse and Mental Health Services Administration (SAMHSA) and is conducted by RTI International in Research Triangle Park, North Carolina.[3] Exhibit 2.1 shows the organizational structure for the 2012 MHSS, including each operational area and task leader at RTI International.

This chapter provides a general description of the activities performed under each MHSS operational area. Overall project management for the MHSS is described in Section 2.1. For each MHSS operational area, when possible, the task leader for the NSDUH main study led the corresponding MHSS task. This management structure provided efficiency gains for the MHSS by taking advantage of the infrastructure, best practices, and staff experience from the main study. However, for some MHSS operational areas, such as data collection, it was beneficial to have a separate task leader focused on managing the completion of MHSS activities because the work for these tasks differed greatly from the main study activities. Unless specified below, the MHSS management structure and general activities completed under each operational area remained the same from 2008 to 2012.

Exhibit 2.1 MHSS Organizational Structure

MHSS = Mental Health Surveillance Study.

[3] RTI International is a trade name of Research Triangle Institute.

2.1 Project Management

The NSDUH Project Director and Associate Project Directors oversaw the schedule, planning, and implementation of major MHSS activities and worked in collaboration with SAMHSA and MHSS task leaders to develop solutions to any anticipated challenges or problems encountered. The mental health task leader oversaw and provided expertise and guidance on clinical activities for the MHSS data collection. For example, this task leader directly supervised the clinical supervisors (CSs) and provided supervision to clinical interviewers (CIs) on challenging cases, distressed respondent cases, and quality control (see Sections 3.2 and 3.3 for more information on CS and CI roles). The mental health task leader was also responsible for consulting as needed with Dr. Michael First, the lead author of the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002), and making recommendations to SAMHSA for SCID revisions each year. The operations task leader oversaw the day-to-day MHSS project operations, including the schedule, budget, staffing, and implementation of each MHSS operational area described below (and shown in Exhibit 2.1). The operations task leader also directly supervised the data collection managers (DCMs) for the MHSS. (See Section 3.1 for more information on the DCM role.)

2.2 Instrument Assessment and Development

The instrumentation activities for the MHSS included developing the computer-assisted interviewing (CAI) recruitment scripts used at the end of the NSDUH interview and the instrument used for the follow-up clinical interview.

For the CAI recruitment scripts, each NSDUH interview respondent who was eligible and selected for recruitment for the MHSS was asked at the end of the interview to participate in the follow-up clinical interview. As shown in Appendix A, the NSDUH interview included a series of carefully scripted recruitment screens to guide the NSDUH field interviewer (FI) through the recruitment steps for the MHSS. Because these recruitment scripts were part of the NSDUH interview, this part of the instrument assessment and development task for the MHSS was led by the same instrument assessment and development task leader who led this task on the NSDUH main study. This task leader, with assistance from other NSDUH survey methodologists, developed the specifications for these recruitment scripts and then tested them once approved by SAMHSA and programmed in the CAI instrument.

Respondents who agreed to the follow-up interview were administered the paper-and-pencil SCID by trained CIs over the telephone approximately 2 to 4 weeks after the NSDUH interview. As discussed in Section 4.2.2, RTI modified the SCID in 2007 for the MHSS to assess past 12-month mental health disorders (versus past 30-day mental health disorders) and functioning via telephone interview by a trained CI. The mental health task leader consulted with and obtained approval from Dr. Michael First and SAMHSA on these initial SCID modifications and any revisions to the SCID each year of the MHSS. Dr. First provided training and guidance on the SCID as needed throughout the course of the study.

The instrument assessment and development task leader was also responsible for incorporating the MHSS instruments, informed consent protocol, materials, and procedures into the Office of Management and Budget (OMB) and Institutional Review Board (IRB) packages for the NSDUH. The MHSS task leaders from each operational area provided input and materials

for the OMB and IRB packages as needed. Chapter 4 includes additional information on the MHSS data collection instruments, informed consent procedures, and OMB and IRB packages.

2.3 Sample Design and Selection

MHSS sample design and selection included designing the MHSS sample, monitoring the sample yield, creating design-based weights, and calculating response rates for each annual sample. During each data collection quarter, the sample design and selection task leader worked with the DCMs to monitor the sample yield and make projections for the anticipated number of completes. At the end of each quarter, the task leader adjusted the sampling rates as needed to yield the desired number of interviews for the year. At the conclusion of data collection, design-based weights and final response rates were calculated by the RTI statisticians who also completed these activities for the NSDUH main study. Chapter 3 of the 2012 MHSS design and estimation report (Liao et al., in press) includes more information on the MHSS sample design and selection, and Chapter 7 of this report provides more detail on MHSS response rates.

2.4 Training and Field Preparation

The activities for MHSS training and field preparation included recruiting, training, and certifying the CSs and CIs; developing training materials for CI training sessions and NSDUH new-to-project (NTP) and veteran FI training sessions; and developing and duplicating survey materials.

Under the direction of the mental health task leader, the DCMs and CSs conducted extensive efforts to recruit, hire, and maintain a team of CIs. The DCMs assisted the mental health task leader with the logistics and administrative aspects of recruiting and hiring the CIs, while the CSs assisted with the candidate interviews. The number of CIs recruited, trained, certified, and staffed was approximately 7 for the 2009 and 2010 MHSS with a target sample of 500 interviews, and approximately 21 for the 2008, 2011, and 2012 MHSS with a target sample of 1,500 interviews.

To maintain NSDUH best practices for training and materials development, the task leader for MHSS training and field preparation was the same task leader who oversaw training and materials development on the NSDUH main study. As described in Chapter 5, all CIs were required to complete a 4-day, in-person training session with a nonclinical agenda detailing administrative tasks and a clinical agenda focusing on SCID administration and distressed respondent protocols. Under the planning and leadership of the training and field preparation task leader, the mental health task leader and CSs developed and trained CIs on the clinical material, and the DCMs and survey specialists developed and trained CIs on administrative activities associated with MHSS data collection. Once training ended, each CI was required to complete and pass a certification with volunteer respondents on the phone before contacting MHSS sample participants to conduct follow-up interviews. The DCMs assisted with the logistics of these certification interviews, including recruitment, screening, and scheduling. Each certification interview was carefully evaluated by the mental health task leader and/or a CS to determine whether the CI trainee properly administered the instrument. Each CI had three opportunities to complete and pass certification or face termination.

In addition to the CI training and certification materials, the MHSS training development activities included developing training materials for main study NTP and veteran FIs, including self-study materials and in-person training. The training and field preparation task leader and staff prepared the training materials on FI-specific protocols and procedures for recruiting respondents for the follow-up clinical interviews. These procedures included following the scripts provided at the end of the NSDUH interview to recruit respondents, providing the follow-up study description for informed consent, giving the $30 incentive to respondents who agreed to complete the follow-up interview, and collecting adequate respondent contact information for CI follow-up.

The training and field preparation task leader and staff also developed and duplicated the materials needed for MHSS recruitment and data collection, including the follow-up study description, reminder card, interview payment receipt, and the MHSS SCID booklet. More information is provided on the recruitment and staffing for MHSS data collection in Chapter 3, the MHSS data collection materials in Chapter 4, and the MHSS training in Chapter 5.

2.5 Data Collection

The main activity for the data collection task was to complete 1,500 follow-up clinical interviews for the 2008, 2011, and 2012 MHSS and 500 follow-up clinical interviews for the 2009 and 2010 MHSS. As described in Chapter 3, one to two DCMs managed production through regular e-mail and weekly conference calls with the CIs and reports provided on the Web-based NSDUH Case Management System (CMS). The DCMs assigned MHSS cases, monitored the record of calls, managed production and costs, sent unable-to-contact letters to unresponsive sample participants, and managed administrative tasks for the CIs (e.g., approving time sheets and expense reports). The DCM assigned individual cases to a specific CI, taking time zone considerations and availability into account. The CI then completed the follow-up clinical interview as soon as feasible after the NSDUH interview was completed, with contact attempts beginning within 24 hours of receiving a case. Each follow-up interview was required to be completed within 4 weeks of the date of the NSDUH interview. During each MHSS interview, the CI completed the SCID on paper and audio-recorded the interview (with respondent permission).

Documenting the CIs' performance and providing feedback to the CIs were integral to the MHSS data collection task and the clinical supervision process. As discussed in Chapter 8, the mental health task leader and CSs conducted these quality control activities for data collection. Within 48 hours of completing the interview, the CI uploaded the audio file to the CMS and edited and shipped the paper SCID to RTI for proper handling, keying, and analysis. The CSs reviewed all SCID booklets item by item, compared the notes provided by the CI with the diagnostic rating, and listened to the accompanying audio files, as needed, to ensure confidence in the data. The mental health task leader and CSs reviewed the full audio recordings for a randomly selected 10 percent of the clinical interviews. The SCID data were also checked electronically for internal coding consistency. The CSs provided feedback to the CIs for each full review conducted. The mental health task leader and CSs also reviewed and provided guidance on all distressed respondent cases.

To further ensure the quality of the data being collected in the clinical interviews, interrater reliability exercises were conducted at the end of each calendar quarter. For these

exercises, the mental health task leader, CSs, and CIs listened to the audio file from a selected interview (available through the Web-based CMS) and independently rated the assessed symptoms (e.g., present, absent, subthreshold, or not enough information) and disorders (e.g., present, absent, or not enough information). The mental health task leader and CSs used these exercises to estimate interrater reliability, to evaluate the diagnostic skills of the CIs, and to provide retraining for the CIs to reduce error in data collected in future quarters. More information on MHSS data collection activities is provided in Chapter 6 of this report.

2.6 Data Management

MHSS data management provided for software and system development and support for CIs, CSs, and DCMs. Because the MHSS used similar software, hardware, and CMS functions and reports as the NSDUH main study, this MHSS task was led by the same data management task leader who led this task in the NSDUH main study. MHSS data management included initial and ongoing development and support for the computer systems used to automate and manage MHSS data collection and workflow. The specific systems developed and supported by NSDUH programmers for the MHSS were

- laptop-based MHSS CMS for CIs, preconfigured with encryption software and custom software to automate the secure upload of audio files using the encryption facilities of the HTTPS protocol;
- Web-based MHSS CMS for assignment and transfer of MHSS cases, management of MHSS workflow, and review of MHSS audio files;
- MHSS data transmission system, including server-side Web services, to allow bidirectional transmission of MHSS data and audio files between RTI and MHSS CI laptops;
- MHSS daily and weekly automatic report generation software;
- interactive MHSS job scheduling and background data processing; and
- MHSS laptop software to integrate Skype voice over Internet protocol telephone services into the MHSS laptop configuration using the Skype application programming interface for .NET.

In addition to those activities, the data management task included daily phone and e-mail support for CIs, CSs, and DCMs. A second-tier technical support team provided this support.

2.7 Data Analysis and Reporting

The primary objective of the MHSS analysis is to produce annual national estimates of serious mental illness prevalence that have sound psychometric properties, that are accurate, and that use similar methodologies, so it is possible to examine trends over time. Secondary objectives include predicting other categories of mental illness, such as any mental illness, mild (or low) mental illness, and moderate mental illness. Under this operational area, estimation methodology was developed by RTI statisticians and SAMHSA to produce annual estimates of mental illness.

In addition to creating a method for estimating mental illness, methodological studies and analytic reports were completed as part of the 2011 and 2012 MHSS. Each methodological or analytic study was led by and staffed with substantive, methodological, statistical, and survey data collection experts based on the mental health study topic. To closely monitor the status and progress of these studies, a research epidemiologist was responsible for the overall management (i.e., staffing, quality, schedule, and budget) for all MHSS analytic studies, and a research methodologist was responsible for the overall management for all MHSS methodological studies. This management structure enabled these task leaders to focus on the completion of these special methods and analytic studies while the overall task leader for MHSS analysis focused on modeling analyses.

Finally, under the leadership of the same operational directors responsible for the national reports and data tables on the NSDUH main study, RTI provided data tables, technical appendices, graphs, and other pictorial displays as requested by SAMHSA for the NSDUH mental health findings report.

2.8 Data Files and Documentation

The main activity for the MHSS data files and documentation task was to produce a MHSS analytic data file and codebook each MHSS year. Because the MHSS data file and codebook activities were similar to the NSDUH main study, this MHSS task was led by the same task leader as the NSDUH main study. The master database was initiated with the raw, transmitted CAI data from the NSDUH interview. The SCID data were keyed and edited by data clerks and merged into the master database. As data processing proceeded and data items became available after editing, coding, imputation, and weighting procedures, they were merged onto the master database. As files were needed either for internal use or for delivery to SAMHSA, the requested data were extracted to working data files, and corresponding codebook information was generated under program control.

For each MHSS year, a preliminary weighted data file was prepared from the master database. This file was delivered to SAMHSA on a CD in PC SAS® format along with a codebook containing frequencies of data items. These data included all edited, imputed, and recoded analysis variables; weights; estimation variables; and variables used to create the population estimates and associated tables and all special tables.

Once all variable creation was completed, a comprehensive data file was extracted from the master database and a codebook was generated. In addition to the edited variables, imputation-revised variables, recodes, weighting and estimation variables, and selected raw data items were included as requested by SAMHSA. The 2008-2012 Adult Clinical Interview Data File was delivered in PC SAS format to SAMHSA on a CD that also included the electronic version of the file's codebook. Introductory text, selected tables, appendices of codes, and further details were added to the system-generated codebook to document this file.

3. Data Collection Staffing

This chapter describes the staffing for Mental Health Surveillance Study (MHSS) data collection, including a description of the data collection manager (DCM), clinical supervisor (CS), and clinical interviewer (CI) positions and responsibilities, as well as any problems encountered with staffing these positions for the MHSS. More information is provided on MHSS training, data collection, and quality control activities in Chapters 5, 6, and 8, respectively.

3.1 Data Collection Managers

Depending on DCM workload and availability, one or two DCMs monitored MHSS field operations and data collection staffing, production, and costs. Over the course of the study, six different DCMs worked on the MHSS. The DCMs reported to the National Survey on Drug Use and Health (NSDUH) Associate Project Director overseeing MHSS operational activities. All DCMs were bachelor's- or master's-level survey managers who also worked on the NSDUH main study. DCMs were responsible for

- assigning follow-up clinical interview cases to the CIs based on respondent and CI availability;
- monitoring the completion of clinical interviews, including reviewing the record of calls, providing guidance to CIs on contact attempts, and directing CIs to assign final disposition codes to clinical interview cases (e.g., codes assigned that indicated interview complete, unable to contact, refusal);
- sending unable-to-contact letters to remind unresponsive MHSS sample participants to contact the DCM and schedule an interview appointment;
- confirming interview audio files were uploaded to the Web-based Case Management System;
- confirming the receipt, editing, and keying of completed Structured Clinical Interviews for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NPs or SCIDs; First, Spitzer, Gibbon, & Williams, 2002) booklets at RTI International;
- producing and reviewing weekly and daily data collection status reports;
- conducting weekly conference calls with CIs to review and discuss case assignment status, schedules, record of calls, challenging cases, and technical problems;
- approving CS and CI weekly timesheets and expense reports;
- assisting with the logistics and administrative aspects of recruiting, hiring, and training CIs;
- assisting with recruitment, screening, and scheduling of certification interviews; and
- monitoring completion of the quarterly interrater reliability (IR) exercises.

3.2 Clinical Supervisors

Depending on the MHSS sample size each year, two to four CSs, all of whom had doctoral-level training in psychology, closely monitored the MHSS data for quality and accuracy. As described in Section 5.5, the CSs were licensed psychologists who had been trained and certified in SCID administration by the SCID's developer, Dr. Michael First. The CSs also received refresher training by the developer each year from 2008 to 2011. The CSs reported to the mental health task leader, who oversaw and provided expertise and guidance on clinical activities for the MHSS data collection. The CSs' responsibilities included

- screening, interviewing, and selecting highly qualified CIs;
- developing and revising clinical materials for and conducting new-to-project and veteran CI training sessions and certification interviews;
- developing and revising certification and production versions of the SCID booklet, including the Distressed Respondent Protocol (DRP);
- developing and revising manuals for editing, keying, and coding/recoding SCID data;
- reviewing and editing completed SCID booklets and audio files for quality and accuracy;
- providing individual supervision and feedback to CIs on the SCID, clinical interviewing issues, challenging cases, and distressed respondents;
- monitoring the CIs' quality of work when administering the SCID and alerting the DCMs of any concerns (see Section 8.1 for more information); and
- conducting quarterly interrater reliability (IR) exercises, which included selecting an interview for review; collecting, scoring, and analyzing the CIs' and CSs' IR data; reviewing the results of the exercise during conference calls with the CIs, CSs, and staff from the Substance Abuse and Mental Health Services Administration (SAMHSA); and submitting a quarterly report to SAMHSA summarizing the IR results.

3.3 Clinical Interviewers

As shown in Table 3.1, 6 to 23 CIs collected SCID data for each year of the MHSS. The number of CIs needed for the MHSS varied by the target number of completed interviews per year, which was 1,500 in 2008, 2011, and 2012, and 500 in 2009 and 2010. On average, 7 CIs per 500 interviews were staffed for the MHSS.

Table 3.1 Number of Clinical Interviewers Staffed for the 2008-2012 MHSS

MHSS	Target Interviews (#)	CIs at Beginning of Year (#)	Replacement CIs Certified (#)	Total CIs (#)	CI Attrition (#)	CIs at End of Year (#)	CI Attrition (%)
2008	1,500	22	3	25	3	22	12
2009	500	8	0	8	2	6	25
2010	500	6	2	8	1	7	13
2011	1,500	21	9	30	7	23	27
2012	1,500	23	0	23	2	21	9

CI = clinical interviewer; MHSS = Mental Health Surveillance Study.

When hiring for the 2008 MHSS, the necessary CI credentials included having a master's or doctoral degree in clinical or counseling psychology, a medical degree with a specialty in psychiatry, or an advanced degree in a related field such as clinical social work. After the first year of the MHSS data collection, it was determined that CIs with a terminal master's degree were not adequately prepared by their graduate training to meet the data quality standards required for this study. Therefore, starting with recruitment for the CI training session held in December 2009 for the 2010 MHSS, all applicants were required, at a minimum, to be enrolled in their third year of a doctoral-level psychology program at the time of applying for the position. CIs were responsible for

- successfully passing CI training and certification before MHSS data collection;
- contacting respondents and scheduling clinical interviews;
- obtaining informed consent and administering the SCID interviews according to MHSS study protocols, including uploading electronic recordings of the interviews and shipping SCID booklets to RTI for editing and keying;
- attending scheduled individual meetings with CSs for supervision and seeking CS supervision as needed;
- participating in quarterly IR exercises, group conference calls, and re-training exercises;
- attending weekly meetings with the DCM and seeking supervision as needed; and
- following the DRP (see Section 4.2.5) when encountering respondents who reported suicidal ideation, reported homicidal ideation, or appeared upset or agitated during the interview.

Potential MHSS CI applicants were drawn from two main sources: (1) American Psychological Association (APA)–accredited graduate programs in clinical and counseling psychology, professional psychology internships, and postdoctoral training (APA Web directory; American Psychology Postdoctoral and Internship Centers directory), and (2) key professional organizations in psychology, psychiatry, social work, and counseling that emphasized research and stated psychological associations (e.g., APA Web directory, Association for Behavioral and Cognitive Therapies Web directory). Additionally, when CI positions became available, former CIs who had proven to be strong interviewers were contacted and encouraged to apply and share the announcement with others.

For the initial training in 2007, a total of 268 Web-based applications were received between September 1 and October 1. The DCMs screened these applications to ensure that the applicants met the necessary criteria. Of the 268 applicants, 84 percent (n = 225) met the necessary criteria and were considered as possible hires. These 225 applications were reviewed by the mental health task leader, who selected the applicants for an interview with a CS. With a goal of hiring 30 CIs, the MHSS CSs conducted interviews with the top 90 CI applicants, of whom 45 (50 percent) were considered for hiring based on their credentials and skills. The CSs rank-ordered these candidates to create a list of 30 newly hired CIs and a pool of 15 applicants who would serve as backup CIs.

For the final training in 2011, changes were made to the application process to streamline CI recruiting. First, as described previously, the CI criteria had been revised beginning with the recruitment for the 2009 CI training session to include, at a minimum, being enrolled in their third year of doctoral coursework. By eliminating applicants who were trained in a terminal master's degree program, applicants were pre-selected with the highest potential for hiring. Second, applicants who met the minimal criteria on the Web-based application were automatically sent an e-mail inviting them to schedule a screening interview with an outside hiring agency (under subcontract to RTI). Applicants who passed the screening interviews were then reviewed by the mental health task leader, who selected the top applicants for an interview with a CS. Using this process, a total of 552 Web-based applications were received between March 1 and April 24, 2011. Of these, 285 applicants (52 percent) met the necessary criteria and were considered as possible hires. These 285 applicants were invited for a screening interview with the hiring agency. The first 84 applicants who passed the screening interview were reviewed by the mental health task leader, who selected the applicants for an interview with a CS. The CSs conducted interviews with 53 CI applicants, of whom 35 (66 percent) were considered for hiring based on their credentials and skills. The CSs then rank-ordered the candidates to create a list of 12 newly hired CIs and a pool of backup CIs.

Applicants who met the necessary criteria were categorized as "High," "Medium-High," "Medium," "Medium-Low," and "Low" for priority of interviewing, based on their past or present experience with a structured diagnostic interview (e.g., the SCID or equivalent), comparatively recent research experience (e.g., adhering to a research study protocol in the past 10 years), diagnostic assessment experience with Axis I disorders, and geographic location (with lower priority for applicants applying in time zones for which there was a sufficient number of "High" priority applicants). Applicants were categorized as "No Priority" if they admitted to no recent research experience (e.g., during the last 10 years) and if they reported limited experience conducting Axis I assessments. Individuals in this category most often were private practitioners whose primary activity was conducting psychotherapy. The top applicants for each of the three time zones (Eastern, Central, and Mountain/Pacific Standard Time) were invited for telephone interviews with the MHSS CSs; these interviews included administering parts of the SCID during a mock interview. Approximately three times the number of applicants needed for hire were invited for interviews (i.e., 90 interviews to fill 30 slots in 2007). A pool of backup applicants who completed the telephone interviews but did not make the first round of hiring was also created. The applicants on this alternate list agreed to be contacted if additional opportunities for hiring arose on the MHSS.

To ensure adequate coverage during peak times of interview requests, the CIs hired for the study were distributed across the Eastern, Central, and Mountain/Pacific Standard Time zones. Based on experience hiring CIs of this caliber for the National Vietnam Veterans Longitudinal Study and the NSDUH Clinical Validation Study, it was anticipated that between 20 and 30 percent of the CIs hired would not pass the certification standards. Therefore, 30 percent more CIs were generally hired than needed for data collection (e.g., in 2007, 30 CIs were hired and trained when 20 were needed for data collection).

3.4 Problems Encountered

Some challenges were encountered with staffing CIs on the MHSS. First, it was determined that CIs with a terminal master's degree were not as prepared for data collection as

those who were enrolled in doctoral programs. For this reason, beginning with recruitment for the 2009 CI training session for the 2010 MHSS, the minimum requirements of applicants for the CI position was raised from a terminal master's degree to requiring that, at a minimum, applicants be enrolled in their third year of a doctoral program at the time of applying for the position.

Second, CI availability was inconsistent depending or personal or professional demands. The "ideal" CI was often at the point of his/her education and professional development where availability changed from year to year; internships, dissertations, and launching one's career often took priority. Many strong CIs cut back their available hours as other personal or professional demands took precedence. Additionally, the employment needs for the MHSS in subsequent years was uncertain at the beginning of the study. CIs, as a result, were unable to rely on continued employment beyond the single-year contract and made commitments to other endeavors.

3.4.1 CI Turnover

As previously noted, 30 percent more CIs were generally hired than needed for data collection. As shown in Table 3.2, in 2007, 2010, and 2011, approximately 75 percent of the CIs passed certification and, in 2009, 67 percent passed certification. However, at the replacement training in 2008, only 38 percent (3 out of 8) passed certification. As previously discussed, this lower rate of CIs passing certification was the result of having a pool of applicants in 2008 with a lower minimum education requirement. In 2011, the decision was made to overhire and increase the size of the training classes to mitigate the potential risk of staff turnover and costs associated with conducting an additional training at a later time. This larger training session in June 2011 resulted in maintaining the minimum number of CIs needed throughout the 2012 MHSS data collection period.

As reported in Table 3.1, annual attrition rates ranged between about 9 and 27 percent. The average annual attrition rate following certification was 17 percent. Of the 50 CIs who passed certification over the 5 years of the MHSS (including 6 who were re-certified), 11 resigned due to changes in personal commitments, and 4 were terminated due to skill deficiency. Before termination, individual feedback and coaching were provided to improve CIs' performance. A core group of 11 CIs from the initial study year (2008) remained with the project for its duration or returned to the study: 5 CIs remained with the study for the entire 5 years, and 6 CIs who were with the study in 2008 returned in 2010 or 2011 for the remainder of the study. The greatest risk of turnover for any CI was in the initial year following certification.

Table 3.2 MHSS Clinical Interviewer Training and Certification Results

Training Date	Attended Training (#)	Passed Certification (#)	Failed Training or Certification (#)	Pass Rate (%)
November 2007	30	22	8	73
March 2008	8	3	5	38
December 2009	3	2	1	67
October 2010	19	14	5	74
June 2011	12	9	3	75
Total	72	50	22	70

MHSS = Mental Health Surveillance Study.

4. Data Collection Materials

RTI International and the Substance Abuse and Mental Health Services Administration (SAMHSA) staff prepared the necessary interview questionnaire, handbooks, and other survey materials for the Mental Health Surveillance Study (MHSS) each year of data collection. This chapter describes the recruitment materials, interview materials, equipment, handbooks, and Office of Management and Budget (OMB) and Institutional Review Board (IRB) packages used on the MHSS.

4.1 MHSS Recruitment Materials

This section provides a description of the computer-assisted interviewing (CAI) recruitment scripts and materials used by the National Survey on Drug Use and Health (NSDUH) field interviewer (FI) to recruit respondents at the end of the NSDUH interview for the MHSS follow-up clinical interview.

4.1.1 CAI Recruitment Scripts

NSDUH interview respondents who were eligible and selected for recruitment for the MHSS first learned of the MHSS at the conclusion of the NSDUH interview. As shown in Appendix A, the CAI instrument included a series of carefully scripted recruitment screens to guide the FI through the recruitment steps. The scripts first introduced the study, asking the respondent to complete the follow-up interview within the next 2 weeks. For persons unavailable during those 2 weeks, the time period was extended to 4 weeks. Respondents agreeing to participate were asked to provide contact information including first name, phone number, an alternate phone number if available, and best days and times to call. The FI then provided a $30 cash incentive for the MHSS interview and concluded the NSDUH interview.

4.1.2 Respondent Materials

Each NSDUH FI had separate Mental Health Study packets containing the materials needed during the MHSS recruitment process. These materials are provided in Appendix B. When introducing the study, the FI gave the NSDUH respondent a Follow-Up Study Description that contained important information about participation. Those respondents agreeing to the additional interview received a reminder card on which the FI recorded the dates and times the respondent provided for potential contact by phone. When providing the $30 cash incentive for the MHSS interview, the FI signed and gave the respondent a Follow-Up Interview Payment Receipt. All MHSS recruitment materials were printed on blue paper to distinguish them from similar NSDUH materials.

4.2 MHSS Interview Materials

This section describes the materials used by the clinical interviewer (CI) during the follow-up clinical interview.

4.2.1 Introduction and Informed Consent

During a telephone contact, the CI first confirmed the identity of the MHSS respondent using the respondent's first name only and asked several questions about the safety and privacy level of the respondent's current location. As needed, an alternate time was established for an interview. To reference these introduction steps, see Appendix C, Introduction pages 1-2.

For respondents indicating a willingness to continue, the CI covered the necessary elements of informed consent by reading a script (see Appendix C, Introduction page 2) verbatim that detailed the purpose and nature of the study. The CI mentioned the Follow-Up Study Description provided earlier by the FI, asked if there were any questions, and then asked the person if it was okay to continue with the interview.

Next, for willing respondents, the CI read a script to obtain permission to make an electronic audio recording of the interview (see Appendix C, Introduction page 3). The respondent's participation was not contingent upon providing permission to record. See Section 4.3.2 for more information about MHSS audio recordings.

4.2.2 SCID Content

Respondents who agreed to the follow-up interview and who were later contacted by phone were administered the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID, 1/2007 revision; First, Spitzer, Gibbon, & Williams, 2002). Based on the criteria for psychiatric illnesses published in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994), the SCID-I/NP is a semistructured interview that has been widely used in clinical calibration studies such as the National Comorbidity Survey-Replication (Kessler et al., 2004), the National Survey of American Life (Jackson et al., 2004), and the NSDUH substance use disorders reappraisal study (Jordan et al., 2008). The SCID has demonstrated good reliability (Zanarini et al., 2000) and validity (Basco et al., 2000). Studies that compared telephone versus face-to-face administration of the SCID have found good agreement (Crippa et al., 2008; Hajebi et al., 2012; Lee et al., 2008). In 2007, RTI modified the SCID for the MHSS to assess past 12-month mental health disorders (versus past 30-day mental health disorders) and functioning via telephone interview by a trained CI.

Diagnostic modules contained in the MHSS version of the SCID are listed in Table 4.1. The modules for lifetime major depressive episode and lifetime manic episode were included to provide context for understanding whether a past year major depressive episode was experienced as part of a unipolar mood disorder or as a component of a bipolar disorder. The module to assess intermittent explosive disorder was obtained from the (optional) Impulse Control Disorders section of the SCID. This module was not one of the core SCID modules, but the developer made it available as an optional SCID module if investigators were interested in measuring intermittent explosive disorder. Appendix C contains the MHSS SCID content.

Table 4.1 Diagnostic Modules Contained in the MHSS SCID

Mood Disorders	**Past Year Eating Disorders**
Past Year Major Depressive Episode	Anorexia Nervosa
Lifetime Major Depressive Episode	Bulimia Nervosa
Past Year Manic Episode	
Lifetime Manic Episode	**Past Year Impulse Control Disorders**
Dysthymic Disorder	Intermittent Explosive Disorder
Past Year Psychotic Disorders	**Past Year Substance Use Disorders**
Psychotic Screen	Alcohol Abuse
	Alcohol Dependence
Past Year Anxiety Disorders	Non-Alcohol Substance Abuse
Posttraumatic Stress Disorder	Non-Alcohol Substance Dependence
Panic Disorder with and without Agoraphobia	
Agoraphobia without History of Panic Disorder	**Past Year Adjustment Disorders**
Social Phobia	Adjustment Disorder
Specific Phobia	
Obsessive Compulsive Disorder	
Generalized Anxiety Disorder	

MHSS = Mental Health Surveillance Study; SCID = Structured Clinical Interview for DSM-IV.

In addition to the diagnostic modules, the MHSS SCID included the following:

- An open-ended overview module, designed to elicit information about the respondent's diagnostic and treatment history and current status in a way that establishes some level of rapport between the interviewer and the respondent.
- A screener module containing questions for several of the anxiety disorders and eating disorders. Asking these screener questions together in an early part of the interview minimizes the risk of negative response bias (respondents giving "no" answers to speed the interview along when they figure out that "yes" responses typically lead to additional questions).
- A module containing the DSM-IV Axis V Global Assessment of Functioning (GAF) scale. The interviewer was instructed to rate the respondent's period of worst psychological, social, and occupational functioning during the past year.
- A module for documentation of the CI's impressions of the interview situation, including ratings of the respondent's level of privacy, cooperation, and comprehension, as well as the overall validity of the interview data.

Per the standard SCID administration protocol, respondents were read the scripted SCID questions verbatim, in combination with unscripted follow-up questions that the CI tailored to the respondent in order to gather sufficient clinical information to assign a code to the necessary criteria for each symptom. CIs rated each symptom as being absent ("1"), present at a subclinical level ("2"), or present at a clinically significant level ("3"). CIs could also temporarily rate a symptom as having not enough information ("?") before returning to rate the symptom as absent, present at a subclinical level, or present at a clinically significant level. Based on the codes for each symptom and the criteria needed for diagnosis, the CIs then followed the instructions in the SCID, which were to proceed to the next question or skip ahead to the next module.

Following completion of the interview, respondents were read an End of Interview script (see Appendix C, End of Interview page K.4). CIs ensured that respondents had the toll-free number to the National Lifeline Network should they want additional information about mental health services in their area. If the CI felt the respondent was at all in distress or potentially was a danger to himself or herself or others, a Distressed Respondent Protocol (DRP) was followed (see Section 4.2.5).

4.2.2.1 MHSS SCID Changes

The SCID's publishers at Columbia University, Biometrics, provided changes to the SCID that were then incorporated into the MHSS version of the SCID for the start of 2010 MHSS data collection.[4] Many of the changes were related to inconsistent use of "?" and "2" ratings for individual SCID items, which Biometrics found upon close examination of the ratings. These errors were corrected so that the use of "?" and "2" were consistent throughout the SCID. The MHSS SCID also incorporated a correction made by Biometrics to ensure that all criteria were met for the diagnosis of panic disorder. None of these changes impacted the data collected in prior years.

Beginning with the 2010 MHSS, codes were added to capture more information about respondents' mental health, which was already being collected by the CIs during the interview as part of the SCID. Therefore, the SCID questions remained the same while variables were added to code the existing data. These new variables included the following:

- type of mental health treatment (counseling, medication, or both) in the past year;
- presence of psychiatric hospitalizations in the past year and the reason(s);
- names of psychotropic medications taken in the past year and/or currently taken;
- score on the Short Blessed Scale (SBS), if administered; and
- potential disorders not assessed for which at least some symptoms were reported or demonstrated by the respondent during the MHSS interview (e.g., bipolar II disorder, attention-deficit/hyperactivity disorder, Asperger syndrome, antisocial personality disorder).

Data for potential disorders not assessed were coded and specified by both the CI and clinical supervisor (CS) in the MHSS SCID booklet (see Appendix C, Interviewer Debriefing pages X.4-X.5). The presence of any disorders not assessed was determined by the respondent's self-reports and the CI's and CS's behavioral observations. The presence of any disorder not assessed was coded by the CI and CS as "1" ("No") or "3" ("Yes"). If any potential disorders not assessed were coded as present ("3"), then the CI and CS specified the name(s) of the disorder(s) implicated for rule-out, described as follows. These data were keyed into open text fields in the dataset. Any potential disorder not assessed was coded according to the following four categories:

[4] SCID changes were posted to the SCID Web site at http://www.scid4.org/revisions/jan10 html.

- Rule-out Other Axis I Disorder (not assessed in study; coded as "1")
- Rule-out Axis II Disorder – Personality Disorder (not assessed in study; coded as "2")
- Rule-out Axis II Disorder – Other (e.g., mild mental retardation; coded as "3"), or
- Rule-out Axis I Disorder assessed but missed (due to CI or respondent error; coded as "4").

4.2.3 GAF Score

At the end of the interview, the CI assigned a GAF score (see Appendix C, End of Interview pages K.2-K.3) to the respondent. Representing the DSM-IV Axis V assessment, the GAF score ranges from 1 to 100, based on a two-dimensional continuum of (1) psychiatric symptoms—from severe mental illness to superior mental health, and (2) psychosocial functioning—from severe impairment to superior functioning. For example, a GAF score of 5 would indicate very serious psychiatric symptoms and/or grossly impaired functioning, a GAF score of 95 would indicate no symptoms and superior functioning across a wide range of psychosocial and occupational activities, and a GAF score of 50 would indicate serious symptoms or serious impairment in social, occupational, or school functioning. Impairment in functioning due to physical or environmental limitations is not included when assigning a GAF score.

For the MHSS, the CI recorded the GAF score that represented both the respondent's worst symptoms and functioning in the past year. A GAF score is determined by rating both symptoms and functioning according to the GAF scale. If the rating based on either symptoms or functioning is lower than the other, then the lower of the two ratings is used for the GAF score. For example, if a respondent's functioning implicated a GAF score of 50, but the symptoms implicated a GAF score of 30, then the GAF score of 30 would be assigned because it is lower. A respondent would be classified as having serious mental illness if (1) the GAF score was 50 or lower and (2) one or more non-substance use disorders were present in the past year.

4.2.4 Short Blessed Scale

A protocol was developed for encounters with respondents suspected of having problems with basic cognitive functions (such as attention, language production, orientation, language comprehension, and memory), which could lead to invalid data. As part of this Cognitive Impairment Protocol, the CIs were instructed to immediately break off the contact if the respondent showed signs of cognitive impairment during the introduction and informed consent process; however, if they had started the SCID, the CIs were to stop the interview and administer the SBS to respondents. The SBS (see Appendix C, Cognitive Impairment Protocol page CIP.1) is a six-item scale designed to assess cognitive ability according to orientation, memory, and concentration. These questions are provided to CIs in the Cognitive Impairment Protocol section at the back of the SCID booklet. SBS scores indicate the number of errors, ranging from 0 to 28. Errors on the SBS can be indicative of temporary cognitive impairment (e.g., alcohol intoxication, medication side effects) or more long-term cognitive dysfunction (e.g., dementia, head injury).

For 10 or fewer errors on the SBS, CIs were instructed to resume the interview and to note the situation in the interviewer debriefing questions at the end of the SCID. For more than

10 errors, CIs were instructed to break off the interview and to document the situation when they entered the interview status code of "56: Breakoff (Partial Interview)" in the Web-based Case Management System (CMS). CIs were instructed to attempt to complete the interview at another time if the respondent's cognitive impairment was deemed temporary (e.g., under the influence of alcohol or some other mind-altering substance). If completing the SCID at another time, CIs were instructed to review the portions of the SCID completed earlier, to verify accuracy, and to include the reason for the breakoff in the debriefing questions. The results of administering the SBS in the MHSS are provided in Section 7.4.

4.2.5 Distressed Respondent Protocol

CIs were trained to fully assess signs and symptoms of emotional distress by the respondent during the administration of the SCID over the telephone. Respondents' distress can include reports of recent suicidal or homicidal thoughts, plans, or actions, or showing strong feelings of sadness, irritability, or agitation during the SCID administration. If the level of imminent danger was immediate or if the level of distress was such that completing the interview would increase the distress of the respondent, CIs were instructed to terminate the interview immediately and follow the appropriate DRP.

As shown in Appendix C, DRP pages DRP.1-DRP.7, the detailed MHSS DRP covered five different situations according to risk of harm to self or others. As summarized in Table 4.2, DRP Scenario #1 was for respondents reporting recent passive suicidal ideation, which included vague thoughts of suicide in the absence of a plan (risk of self-harm, no imminent danger). DRP Scenario #2 was for respondents reporting recent active suicidal ideation, which included specific plans for suicide and acting on those thoughts (risk of self-harm, possible/definite imminent danger). DRP Scenario #3 was for respondents reporting recent passive homicidal ideation, which included vague thoughts of homicide in the absence of a plan (risk of harm to others, no imminent danger). DRP Scenario #4 was for respondents who reported recent active homicidal ideation, which included specific plans for homicide and acting on those thoughts (risk of harm to others, possible/definite imminent danger). DRP Scenario #5 was for respondents who showed signs of emotional distress during the SCID, such as sadness, irritability, or agitation, in the absence of suicidal or homicidal thoughts (no risk of harm, respondent is agitated or upset).

Table 4.2 Distressed Respondent Protocol Scenarios

Distressed Respondent Protocol Scenario	Risk of Harm	Imminent Danger
Scenario #1: Passive suicidal ideation	Self	No
Scenario #2: Active suicidal ideation	Self	Yes/Maybe
Scenario #3: Passive homicidal ideation	Other(s)	No
Scenario #4: Active homicidal ideation	Other(s)	Yes/Maybe
Scenario #5: Respondent agitated or upset	None	No

To ensure respondents' safety, the DRP was changed in 2011 to incorporate the following:

- If a respondent reported recent active suicidal ideation (Scenario #2), the CI would connect the respondent with emergency mental health services in his or her area using

SAMHSA's online database of inpatient mental health providers before contacting local emergency rescue services (that may be less prepared for handling suicidal respondents).

- If a respondent reported recent active suicidal thoughts (Scenario #2) or recent active homicidal ideation (Scenario #4), the CI would connect the respondent with local emergency rescue services (using a searchable national database of numbers that could be used to directly contact emergency rescue services nearest to the respondent) rather than calling 9-1-1 (which automatically routed calls to the emergency rescue services nearest to the CI, not necessarily the respondent, thereby interfering with a respondent being connected with local emergency rescue services).
- If a respondent reported passive suicidal ideation (Scenario #1), reported passive homicidal ideation (Scenario #3), or became upset or agitated during the interview (Scenario #5), respondents were encouraged to call Lifeline's number directly rather than the CI offering to connect them using three-way calling (which sometimes got disconnected or automatically routed calls to the Lifeline nearest to the CI, thereby interfering with a respondent being referred for local mental health services).

To reduce the likelihood of contacting emergency rescue services prematurely, the DRP was changed for Quarter 2 of 2012 to incorporate the following: If a respondent reported recent active suicidal thoughts (Scenario #2) but denied intentions to harm themselves, the CI would strongly suggest mental health counseling rather than connecting him or her with local emergency rescue services (which may not provide access to mental health services). The updated DRP Booklet provided to CIs in 2012 is included in Appendix D (Addendum 2, page DRP-1). Section 7.3 provides the number of respondents classified as distressed during the 2008-2012 MHSS.

4.2.6 Interviewer Debriefing

Following the conclusion of the interview, the CI completed the Interviewer Debriefing module (see Appendix C, Interviewer Debriefing pages X.1-X.5), which documented the CI's impressions of the interview situation, including ratings of the respondent's level of privacy, cooperation, and comprehension, as well as the overall validity of the interview data. This module also allowed the CI to note "rule-out" disorders that were not assessed but may have impacted the GAF score in the past year. Upon independent review, the clinical supervisor also rated each interview for global validity and noted any potential "rule-out" disorders that were not assessed. Data from interviews with questionable validity were not included in the final dataset.

4.3 CI Equipment

This section describes the equipment and systems used by the CI during the follow-up clinical interview.

4.3.1 Laptop

Although the MHSS interview was conducted using paper questionnaire booklets, CIs used NSDUH-issued Gateway laptops for several tasks, including making an audio recording of the interview and uploading the resulting audio file to RTI, management tasks such as case

assignment and reporting, and staff communication via e-mail. For security purposes, NSDUH laptops require two separate user name and password entries to reach the desktop, with additional user name and password entries necessary to access the specific Web-based CMS used for the study.

4.3.2 Audio Recording

With respondent permission, CIs made an audio recording of the MHSS interview. Between January 2008 and June 2011, CIs connected the laptop to a telephone using a specialized recording control device. CIs had two devices available: one for analog telephone systems and one for digital systems. Due to ongoing issues with the quality of the resulting audio recordings, CIs began using the Internet-based phone provider Skype in June 2011 for the interview contact with respondents. Skype is a voice over Internet protocol that enables peer-to-peer voice connections (i.e., telephone calls) over the public Internet. Skype connections are secured using 256-bit Advanced Encryption Standard encryption to ensure the privacy of the communication channel. CIs used a Skype headset with a microphone that attached easily to the laptop. Calling via Skype allowed a more direct connection, which provided quality recordings and was straightforward to use. In the event that no Internet connection was available at the time of an interview, CIs still had the option to record the interview using one of the recording control devices.

A specialized computer system on the laptop developed specifically for MHSS allowed the CI to select the appropriate case, conduct a test audio recording (to test functionality and quality), and then, with respondent permission, record the interview. Upon conclusion of the respondent contact, the system allowed CIs to securely upload the audio file directly to RTI servers.

4.3.3 Case Management System

CIs accessed the MHSS Web-based CMS to check case assignments, provide updates on each respondent contact via case status reports, and, as needed, prepare and submit reports for any distressed respondent situations encountered. Data collection managers assigned MHSS cases based on CI availability during the best days and times to call indicated by respondents. Cases could be transferred among the interviewing staff.

The CMS displayed case assignment information collected by the NSDUH FI, including first name, phone number, alternate phone number if provided, and best days and times to call, along with any helpful notes the FI may have entered. The sample participant's home State and time zone are also displayed to assist CIs in properly timing their contacts. The status code and any comments for each attempt to contact a case were available so the CI could see the history of the case.

Also through the CMS, RTI staff carefully tracked each case from initial assignment through interview completion including questionnaire receipt and processing.

4.4 MHSS Handbooks

Interviewing staff used detailed handbooks both as a learning tool and a reference to the required procedures and protocols of the study.

4.4.1 MHSS CI Handbook

The MHSS CI Handbook detailed all aspects of an interviewer's work requirements on the MHSS. The MHSS CI Handbook covered the following topics:

- Introduction to the Study,
- Your Role on the Mental Health Study,
- Equipment Use and Care,
- Contacting Respondents,
- Conducting the Interview,
- Documenting Results,
- Administrative Procedures, and
- Quality Control.

The MHSS CI Handbook files, including the final CI Skype User Guide (Addendum 1, page S-1) and the DRP Booklet (Addendum 2, page DRP-1), are provided in Appendix D and are discussed as follows.

This handbook was sent to all newly hired CIs for review prior to attending CI training. It was used throughout the training session and served as a ready reference when questions arose during work.

Overall, MHSS procedures remained consistent between survey years, allowing veteran staff to continue to reference their CI Handbook. MHSS CI Handbooks were distributed to CIs at the following points: during their initial training for the 2008 study, prior to the start of the 2010 survey year, and in October 2010 when preparing for the 2011 study. For 2009 and 2012, any administrative updates to the handbook for the upcoming year were provided via memo to the CIs.

In June 2011, all new and veteran staff received two separate addenda to the existing handbook:

- The CI Skype User Guide provided instructions for establishing and using a Skype account to contact a respondent and record the interview and for uploading the resulting audio file to RTI (as discussed in Section 4.3.2).
- The DRP Booklet provided additional instructions on properly implementing the updated protocol for MHSS. This booklet contained the instructions and scripts to use when speaking with a distressed respondent and, if needed, an emergency care provider; the detailed instructions on how to locate the proper 9-1-1 emergency number for the respondent's location; and details about submitting an incident report.

Following further updates to the DRP in the spring of 2012 (as discussed in Section 4.2.5), all staff received an updated DRP Booklet for reference beginning in Quarter 2 of 2012.

4.4.2 MHSS FI Handbook

The MHSS FI Handbook provided detailed instructions for the FIs' tasks on the MHSS. The MHSS FI Handbook covered the following topics:

- Introduction to the Study,
- Equipment and Materials,
- Initial Interview,
- Obtaining Participation,
- Administrative Procedures, and
- Quality Control.

The MHSS FI Handbook files are provided in Appendix E.

Newly hired new-to-project (NTP) field staff received an MHSS FI Handbook for review prior to attending NTP training. All veteran NSDUH FIs received an MHSS FI Handbook to review and reference prior to the start of each survey year with one exception. For the 2010 MHSS, veteran NSDUH FIs received a brief updates memo along with instructions to continue to use and refer to the 2009 MHSS FI Handbook during 2010.

4.5 OMB and IRB Packages

In order to obtain the necessary approvals and clearances to conduct the MHSS, the MHSS procedures were described in annual OMB and IRB packages submitted for the NSDUH main study.

4.5.1 OMB Package

Text describing the purpose of and justification for the MHSS were added to the NSDUH OMB Supporting Statement. This document described procedures for collecting MHSS data, justification for imposing the additional burden on NSDUH respondents, confidentiality measures for NSDUH data, and procedures for protecting MHSS respondents' data. Details on the process for certifying CIs were included in the OMB Supporting Statement. MHSS respondent materials planned for use in the study were included in the OMB package. The package included the Follow-Up Study Description, Follow-Up SCID Interview Content, Introduction to the Clinical Follow-Up Interview, Follow-Up Interview Payment Receipt, and Confidentiality Pledge signed by CIs staffed on the study.

OMB clearance to conduct the MHSS as a follow-up study to the NSDUH was granted annually. The 2009 OMB package described the plans for both the 2009 and 2010 MHSS, and clearance was received for these 2 survey years.

4.5.2 IRB Package

RTI's IRB Committee met annually to review and approve the NSDUH project. From 2008 to 2012, this process included a review of the MHSS materials and procedures. Additions to the IRB's Request for Approval of Research Protocol described the procedures of the MHSS

and how these procedures would protect human subjects. The IRB granted approval to conduct the MHSS annually.

Per IRB procedures, project staff reported adverse events that occurred throughout the duration of the study to RTI's IRB. An adverse event could occur if a study respondent were to become upset or distressed during the administration of the instrument. When such an event occurred, the CI followed the steps prescribed in the IRB-approved DRP. The administration of this protocol was then reported to the IRB during the end-of-year closeout reports. Cases that involved DRP violations and/or emergency services (DRP Scenario #2 or Scenario #4) were immediately reported to the IRB.

5. Staff Training

This chapter describes the training programs conducted for field interviewers (FIs), clinical interviewers (CIs), and clinical supervisors (CSs) for Mental Health Surveillance Study (MHSS) recruitment, data collection, quality control, and supervision. Training for these staff occurred prior to the start of data collection each year as well as during the year.

5.1 MHSS Recruitment FI Training

This section describes the programs conducted to train veteran FIs and new-to-project (NTP) FIs on the MHSS recruitment process.

5.1.1 Veteran FI Training

Prior to the start of each survey year, the National Survey on Drug Use and Health (NSDUH) veteran FIs continuing on the study completed a multipart training program to review changes for the upcoming year and to refresh their knowledge of important study protocols.

Self-study tasks for the MHSS completed by FIs in November/December included reviewing the FI Handbook for the upcoming year and completing an MHSS iLearning course. For the 2010 survey year, the MHSS topics were included as part of the NSDUH main study Practicing Perfection iLearning course instead of as a separate MHSS course.

iLearning, an electronic multimedia interactive training application, allowed FIs to complete the training course at their own pace on their NSDUH laptop and review portions of the course again as needed. The MHSS course consisted of visual slides with text and graphics, an audio component providing important information and instructions, and an assessment portion to ensure the FI's comprehension of the material presented. Beginning in 2010, the content also included training videos and interactive practice exercises. Once a course was completed and the results were successfully transmitted to RTI International from the FI's laptop, the course assessment results were posted to the Web-based NSDUH Case Management System (CMS) for field supervisor and project management review.

Each year in early January, veteran FIs attended an in-person Team Meeting to review key project protocols and procedures, which included a review of MHSS topics. Any FIs unable to attend the meeting in person received training during a conference call. The MHSS portion of the training included instruction on MHSS recruitment steps, results, and lessons learned from prior study years. The training in 2011 and 2012 also incorporated exercises to practice the recruitment steps and answering respondent questions, respectively.

5.1.2 NTP FI Training

Newly hired NSDUH FIs received an MHSS FI Handbook (see Appendix E) for review prior to attending the NSDUH NTP FI training session. FI trainees were instructed to review the handbook as part of their preparations for training. After learning and practicing the NSDUH main study interviewing protocols, the trainees were introduced to the MHSS during NTP FI training. Trainers reviewed terminology and detailed recruitment steps along with how to use the

MHSS materials. Training topics also covered ways to answer respondent questions about participation in the MHSS. The trainees watched a video demonstrating the recruitment process and completed a paired exercise to practice all relevant recruitment steps.

5.1.3 FI Refresher Training

Each survey year, MHSS training for FIs continued during the year to provide ongoing refresher training. Prior to each quarter of data collection, all active NSDUH FIs completed their MHSS iLearning training before beginning work. The MHSS iLearning course content and assessment questions varied somewhat for the different quarters to provide helpful training to the FIs. As mentioned in Section 5.2.1, the MHSS content for 2010 was included as part of another veteran iLearning course for the main NSDUH study, also completed quarterly.

5.2 NTP CI Training Sessions

Over the course of the MHSS, four NTP CI trainings were held in November 2007, March 2008, October 2010, and June 2011. Two sessions—in November 2007 for the initial 2008 MHSS and in October 2010 for the expanded 2011 MHSS sample—had two classrooms each of CI trainees, whereas the other two attrition training sessions were smaller, with just one classroom. The first session in 2007 took place in Cincinnati, Ohio, while the other sessions were held in the Raleigh, North Carolina, area.

The NTP CI training program consisted of an initial self-study assignment, 4 days of classroom training, and a certification process. Prior to attending training, CI trainees carefully reviewed the CI Handbook (Appendix D) and the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002) questionnaire booklet (Appendix C). They also completed an online Institutional Review Board (IRB) training course, which covered ethics and regulations involving research on human subjects, the role of the IRB, and the role of the interviewer in protecting respondents' rights. Trainees also completed an online course on the Confidential Information Protection and Statistical Efficiency Act (CIPSEA) requirements that are used to protect information collected on the NSDUH. The course also described the role of the Office of Management and Budget in providing oversight and designating statistical agencies under CIPSEA.

As part of their pre-training work for the October 2010 and June 2011 sessions, CI trainees also listened to an audio recording of an MHSS clinical interview and completed an Interrater Reliability (IR) Exercise form, which they provided to trainers upon arrival at training.

5.2.1 Training Staff

Classroom training staff included a site leader, a technical support staff member, and training team members.

The site leader at each training site managed all registration and fingerprinting activities, hotel relations, and logistics. The site leader coordinated the evaluation of CI trainee performance, working with trainers to resolve any problems, and also reported the status of training to management and supervisory staff each evening. The site leader role was filled by an experienced NSDUH manager with a leadership role on MHSS.

The technical support staff member provided assistance on all technical issues, including the proper functioning of all equipment and programs. Other duties included supervising training equipment setup and the initialization and distribution of CI computer equipment.

Each classroom was taught by a training team consisting of various MHSS staff working closely together to train on tasks in their primary areas of focus. Data collection managers (DCMs) and members of the NSDUH training and materials development team led the training sections on contacting the respondent, using the equipment for audio recording, e-mail, reporting, and administrative tasks. CSs with in-depth knowledge of the SCID instructed CI trainees in the proper administration of the SCID. The mental health task leader (or lead CS) also provided detailed training on handling unusual situations, such as cognitive impairment and proper implementation of the Distressed Respondent Protocol (DRP).

Training teams used a near-verbatim training guide for the non-SCID training topics, along with detailed discussion slides and exercises to review the content of the various SCID modules. During training, each CI trainee received a training workbook with exercises, instructions, and copies of the SCID presentation slides. Other training materials used for each session included mock interview scripts, SCID booklets, and various administrative forms such as training expense reports and a formal calling card agreement for project use.

5.2.2 Content of NTP CI Training Sessions

Over the course of the 5 years of the MHSS, the content of the 4-day NTP CI training session remained consistent with some minor shifting of topics between the training days to improve coverage and timing. The training program content is described below by day. Appendix F provides the training agenda for the June 2011 NTP CI training session.

5.2.2.1 Day 1

Training classes began with an introduction to both the NSDUH and the MHSS. Next, trainees reviewed the responsibilities of the CI position and were instructed on the proper use of the laptop computer hardware and the systems used for case assignment and reporting, recording interviews, and e-mail. CI trainees learned about the procedures for contacting MHSS cases, including the contact information provided for each case in the Web-based CMS, the window of time for making contacts and completing the clinical interview, and the introduction and informed consent process once contact was made. CI trainees then practiced answering respondent questions about the study, referring as needed to sample questions and responses found in the CI Handbook. Management reporting tasks were also covered on Day 1, including entering result codes and the proper handling and distribution of the interview materials. Day 1 ended with training on administrative procedures, including the proper completion of production, time, and expense reports.

CI trainees were assigned homework to practice recording and uploading audio files. Trainers were available for 2 hours in the evening via phone for a "virtual" clinical interviewer lab (CI Lab) session to provide support and answer any equipment-related questions. CI trainees were expected to complete the homework assignment and to review the SCID and interviewing topics in the CI Handbook as preparation for Day 2.

5.2.2.2 Day 2

Day 2 began with an overview of the SCID, followed by reviews of various SCID modules. For each module, trainers explained key concepts of the module, highlighting any issues or differences specifically applicable to the MHSS. Following the detailed review, trainers used a prepared script to lead the CI trainees through an exercise on the content of that module. Day 2 concluded with instructions on the proper use of the Global Assessment of Functioning (GAF) scale. CI trainees completed several GAF exercises during class and were assigned additional GAF exercises for homework. All CI trainees were invited to attend an evening CI Lab to have any questions answered and work one-on-one with a trainer as needed.

5.2.2.3 Day 3

On Day 3, three of the four NTP CI training sessions began with a review of the GAF homework. For the June 2011 session, CI trainees learned about handling unusual situations on the morning of Day 3 and reviewed the GAF homework on the morning of Day 4.

Day 3 training focused on gaining experience and confidence by conducting several exercises on the clinical interviewing process. After learning about the IR exercises conducted quarterly on the MHSS (in which CIs listen to an interview while independently rating the responses and then discuss their ratings as a group), CI trainees completed an IR exercise in class. After a review of instructions in the SCID booklet for the end of the interview and debriefing tasks, CI trainees worked in pairs to practice properly administering the SCID. All CI trainees were again invited to attend an evening CI Lab with trainers available to answer questions or for additional practice.

5.2.2.4 Day 4

The key training objective of Day 4 was to practice every step of the clinical interview process, from case assignment to properly conducting the interview, through reporting and shipment of the completed SCID booklet. Following this paired exercise, CI trainees learned how to handle unusual situations. As previously mentioned, CI trainees completed this training topic on Day 3 at the June 2011 session and began Day 4 with a review of GAF homework, and then completed the paired practice exercise.

To wrap up the training session, CI trainees received instructions and appointment information for their initial certification interview and completed a training evaluation to provide feedback on the completed training session.

5.2.3 Certifications

To ensure CI trainees could complete MHSS interviews following protocol, they were required to successfully complete a certification protocol following completion of the NTP CI training.

For a certification, CI trainees completed the entire clinical interview process, including administering and recording the SCID interview, uploading the audio file to NSDUH's Web-based CMS, and shipping the completed hard copy of the SCID to RTI. Upon receipt, each SCID interview was carefully evaluated by CSs to determine whether the CI trainee properly

administered the instrument. Feedback was provided to the CI trainee, and, as needed, another certification interview was scheduled. CI trainees had up to three attempts to pass certification.

The respondents for these certification interviews were recruited from community mental health and substance abuse treatment centers and were pre-screened to ensure they met the necessary requirements (i.e., 18 years or older, received mental health treatment in the past 12 months, and able to demonstrate comprehension of informed consent). Due to the amount of psychopathy in a clinical sample recruited from a community-based setting, certification interviews lasted about 2 hours, which is longer than the average MHSS interview (lasting about 72 minutes, as described in Section 6.2.1). Volunteer respondents were given $40 for participating in the telephone interview.

Table 3.2 in Chapter 3 provides the MHSS CI training and certification results.

5.3 Veteran CI Training Sessions

Prior to the start of a new survey year and during the survey year as needed, the experienced, or veteran, MHSS CIs participated in training conference calls to review changes and to further refresh and enhance their interviewing skills and knowledge of the MHSS. In addition to veteran training conference calls, one veteran CI training occurred in person in conjunction with the NTP CI Training session held in October 2010.

5.3.1 Staffing

Veteran CI training conference calls were led by various MHSS project staff. A DCM reviewed topics relevant to his or her project tasks, such as production results, administrative details, and goals and expectations for the coming year. When necessary, DCMs also trained on equipment and computer system-related topics. Based on common problems noted during SCID reviews and discussions with CIs, CSs presented information regarding the proper administration of the SCID for the MHSS, handling unusual situations (e.g., uncooperative, upset, or cognitively impaired respondents), and properly implementing the DRP. The trainers used a detailed training agenda as a training guide and schedule for each conference call.

For the veteran CI in-person training held in October 2010, the DCMs and CSs who were NTP CI trainers also served as training staff for the veteran CIs.

5.3.2 Content of Veteran CI Training Sessions

The content of the veteran CI training sessions is described below for the annual sessions conducted prior to each survey year as well as the trainings conducted during the year.

5.3.2.1 Annual Veteran CI Training

In December each year prior to beginning a new survey year, veteran CIs (continuing on the MHSS) completed both independent and group training to review changes for the upcoming year and to refresh their knowledge of important study protocols. Before the group training conference call, CIs were required to complete a few self-study tasks, including completing the annual CIPSEA training course online and reviewing the CI Handbook and any changes for the

upcoming year. With the exception of December 2011, CIs also completed exercises on GAF scale scoring each year as part of their self-study efforts.

The annual veteran CI group conference calls covered the following topics:

- Welcome and Introductions;
- Management, Production, and Data Collection Results and Tips;
- Goals and Expectations;
- DRP Review and Practice Exercises;
- SCID Review and Issue Resolution;
- Review of GAF Exercises (when included);
- Handling Challenging or Unusual Interview Situations;
- Question-and-Answer Session; and
- Wrap-Up.

5.3.2.2 As-Needed Veteran Training Sessions

Twice during the MHSS, changes to protocols required a special veteran CI training during the year. Each time, the CIs received updated materials, performed self-study tasks, and participated in a group conference call. For more information on the protocol changes necessitating these trainings, refer to Sections 4.2.5 and 4.3.2. The following two special CI trainings were held:

- June 2011: CIs reviewed a Skype User Guide, updated their laptops, and established a Skype account. CIs also read an updated DRP Booklet. Changes were reviewed, and all questions were answered during the group call.
- March 2012: CIs received an updated DRP Booklet and completed a related exercise. The changes were reviewed during the regularly scheduled IR call prior to implementation for Quarter 2 of 2012.

5.3.2.3 Rehired Veteran CI Training

At the end of 2009, three CIs who worked on the MHSS in 2008 but not in 2009 were rehired to work again in 2010. Due to prior experience, their refresher training was conducted through detailed self-study and participation in two separate conference calls. The first call, led by the DCMs, reviewed the equipment and administrative tasks as well as MHSS protocols for contacting respondents, reporting results, and submitting the audio file and completed SCID to RTI. The second call, led by CSs, reviewed important details for properly conducting an MHSS interview and handling any unusual respondent situations that may occur. Following these calls, the CI trainees completed the certification process. The CIs also participated in the annual veteran training call held later in December 2009. Two of the three CIs passed certification and worked on the MHSS in 2010.

5.3.2.4 2010 In-Person Veteran CI Training

To coincide with the in-person NTP CI training session held in October 2010, veteran CIs attended the SCID-related content training (Days 2 and 3) of this session as a refresher on clinical interview protocols. These veteran CIs listened as CSs presented information about proper SCID interviewing techniques and assisted the CS trainers by serving as respondents for scripted exercises. Additionally, the veteran CIs met one evening to be trained by the DCMs on detailed administrative topics being implemented in 2011.

5.4 IR Exercises

To ensure the quality of the data being collected in the clinical interviews, IR exercises were conducted at the end of each calendar quarter. The IR exercises, which included group conference calls to calibrate CIs' ratings with the CSs' consensus ratings, were designed to estimate IR to evaluate the diagnostic skills of the CIs and to provide retraining for the CIs to reduce error in data collected in future quarters.

Clinical interviews used for the IR exercises were selected from the pool of recorded interviews conducted during the calendar year. For Quarters 1 and 2 of 2008, the focus of these end-of-quarter exercises was evaluation and retraining using more complex cases, while the focus shifted to obtaining an estimate of typical IR for the remainder of the MHSS using typical cases. Each IR exercise provided opportunities to evaluate the CIs' diagnostic skills and retrain CIs as needed in rating symptoms, diagnoses, and GAF scores.

At the end of each quarter, CSs and CIs listened to the audio file from the selected interview (available through the CMS) and independently rated the assessed symptoms (e.g., present, absent, subthreshold, or not enough information) and disorders (e.g., present, absent, or not enough information). Having reviewed and rated the interview independently, the CSs met as a group and developed consensus ratings of the symptoms and disorders that became the key for comparison with the CI ratings. CIs had 5-7 days to complete the SCID ratings. CIs used a SCID to record their notes and ratings, then they transferred their ratings to the IR Exercise Rating Form.

Agreement between the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994) diagnoses rated by the CSs and the CIs was quantified using Cohen's kappa (Cohen, 1960), a reliability statistic that corrects for chance agreement. A kappa of 0.61 or higher is considered substantial reliability (Landis & Koch, 1977). For each symptom, a reliability coefficient was computed across the SCID rating categories: not enough information (?), absent (1), subclinical (2), and clinically present (3). For each CI, the total percentage of agreement between his or her ratings and the CS consensus ratings across all symptoms and disorders were calculated. The agreement was examined between the CS consensus and CIs' ratings of GAF, the presence of any mental disorder (unrelated to substance use), and the presence of a serious mental illness (SMI)—defined as any mental disorder (unrelated to substance use) and a GAF of 50 or below.

IR exercises were conducted at the end of every quarter except one (Quarter 2, 2009 MHSS, when all CIs were trained/retrained), for a total of 19 exercises. As shown in Table 5.1, participation rates were high, with 86 percent to 100 percent of CIs who collected data during the

quarter participating. CIs who were terminated, who left for personal reasons, or who took a temporary leave of absence did not participate in the end-of-quarter exercise. In the final quarter of the study (Quarter 4, 2012 MHSS), two CIs who were actively employed did not participate due to personal reasons.

The interviews selected for the exercises varied: 11 of the 19 cases (58 percent) were categorized as SMI; that is, the respondents were diagnosed with at least one mental disorder not related to substance use and had a GAF score at or lower than 50. The average GAF score across all IR exercises was 50.7, with a range of 20 to 85.

In many cases, the IR interview was chosen because the GAF score was close to the cut point for classifying the case as having SMI and was therefore more challenging for the CIs to score. The GAF scale, which represented the CI's clinical judgment of the respondent's overall level of "functioning on a hypothetical continuum of mental health-illness," was based on both functioning and symptom severity. To rate the GAF according to the respondent's worst functioning in the past year, the lower of the two scores was used for anchoring the GAF score. As discussed below, problems with reliability in GAF scale ratings (and classification of SMI) in the MHSS appeared to occur mostly from CIs overestimating the seriousness of symptoms, including suicidal ideation. The GAF scale has since been dropped from the DSM-5 due to its conceptual lack of clarity (i.e., including symptoms, suicide risk, and disabilities in the descriptors) and questionable psychometric properties (American Psychiatric Association, 2013, p. 16). As ongoing problems occurred with the reliability of GAF scale ratings, several steps were taken to improve consistency within the MHSS team of CIs and CSs. First, quarterly IR data collection was conducted with group conference calls to discuss and "calibrate" the team's IR ratings. Second, all CI and CS training sessions incorporated GAF lectures and exercises. Third, the CIs and CSs were provided with training and/or supervision by Dr. Michael First (the SCID developer). Despite these efforts, the agreement on the GAF scale ratings remained unreliable.

Over the course of the MHSS, there was high IR agreement between the CIs' and CSs' consensus ratings. Symptom-level agreement was calculated using the total number of variables in the clinical interview as the denominator, and without penalty for omissions and commissions from earlier errors within the same disorder. As shown in Table 5.1, symptom-level agreement improved and was consistently high after the first year, averaging 96 percent. There was also consistently high agreement for having 1 or more mental disorders, with an average of 99 percent. CIs' kappa scores steadily improved each year, with 100 percent of CIs having high kappa scores for the last seven quarters of the study. Less consistent agreement was found for GAF (0 to 95 percent) and SMI (0 to 100 percent). The majority of error for GAF and SMI ratings was due to CIs' GAF ratings that were lower than the CSs' consensus GAF ratings. Discussions during the IR calibration conference calls indicated that CIs' biases toward lower GAF scores were due to overestimations of the severity of the respondent's symptoms (e.g., the seriousness of suicidal thoughts, the danger of driving while impaired by alcohol).

Following each IR exercise, the CSs and CIs participated in group conference calls to calibrate the CIs' ratings to the CSs' consensus ratings. All CIs were encouraged to attend the group calibration conference calls, and over the course of the study, there was 82 percent participation in these calls. Individual make-up calls were held with each CI who could not attend the group conference calls. CIs whose scores were in the lower 33 percent of the

distribution were retrained by the CSs. In all cases, the CIs were able to demonstrate understanding of the content of retraining and, therefore, were not considered for being placed on probation. Using the pre-established cutoff score of 70 percent or lower, no CI was ever considered for termination during the course of the MHSS because of his or her performance on the IR exercises.

Prior to the IR conference calls, CIs were sent an electronic copy of their scored IR rating sheet that listed their ratings compared with the CS consensus ratings and their overall percentages of agreement. The calibration conference calls lasted approximately 1 hour and included a review of all symptoms that were assessed, with particular attention devoted to symptoms for which there was disagreement. Review of each problematic symptom started with a brief rationale for the CS consensus ratings, followed by comments and questions from the CIs and retraining for common sources of error (e.g., inadequate probing, differences between clinical and subclinical symptom ratings). The mental health task leader and CSs also used examples from the stimulus interview to underscore key information that was needed to confidently code a symptom.

Table 5.1 Interrater Reliability Exercise Performance

			CI Performance					Case Description	
Year	**Q**	**CIs Participating/ Employed (%)**	**Agreement: Symptoms (%)**	**Agreement: ≥ 1 Diagnosis (%)**	**Agreement: GAF Decile (%)**	**Agreement: SMI (%)**	**CIs with Kappa ≥ 0.61 (%)**	**CS GAF**	**SMI**
2008	**1**	22/22 (100)	90	100	67	95	100	43	SMI
	2	22/25 (88)	80	100	18	100	23	20	SMI
	3	21/23 (91)	91	100	81	100	100	85	Non-SMI
	4	21/21 (100)	93	100	95	100	100	52	Non-SMI
2009	**1**	8/8 (100)	97	100	88	88	100	58	Non-SMI
	2	--	--	--	--	--	--	--	--
	3	7/7 (100)	97	86	43	100	71	65	Non-SMI
	4	6/7 (86)	99	100	83	100	100	45	SMI
2010	**1**	8/8 (100)	99	100	25	25	100	49	SMI
	2	8/8 (100)	96	100	0	0	100	49	SMI
	3	8/8 (100)	97	100	75	100	100	45	SMI
	4	8/8 (100)	99	100	38	100	100	32	SMI
2011	**1**	20/21 (95)	96	100	60	60	95	50	SMI
	2	18/20 (90)	98	100	89	89	100	53	Non-SMI
	3	23/26 (88)	98	100	91	96	100	45	SMI
	4	23/24 (96)	96	96	57	66	100	53	Non-SMI
2012	**1**	22/23 (96)	98	100	68	68	100	53	Non-SMI
	2	22/22 (100)	98	100	55	55	100	50	SMI
	3	21/21 (100)	98	95	90	100	100	66	Non-SMI
	4	19/21 (90)	99	100	63	63	100	50	SMI
Average		96	96	99	62	78	94	51	--

-- Not available.

CI = clinical interviewer; CS = clinical supervisor; GAF = Global Assessment of Functioning; Q = Quarter; SMI = serious mental illness.

5.5 CS Training

In June 2007, four CSs were trained by the SCID's developer, Dr. Michael First, who was hired as an expert consultant for the MHSS. The CS training included an overview of

interviewing with the SCID, followed by reviews of various SCID modules, including key concepts of the module and highlighting any differences for the MHSS version of the SCID. Dr. First also trained the CSs in using the GAF scale. NSDUH trainers taught the CSs how to perform the technical and administrative procedures using the MHSS laptops. During the training, the CSs completed several practice exercises with the SCID, GAF scale, laptops, technical procedures, and administrative tasks.

To ensure that CSs could provide proper supervision of MHSS interviews, each CS completed the same rigorous certification process described in Section 5.2.3 for the CIs. Each CS certification interview was carefully evaluated by Dr. First, who determined whether the CS properly administered the instrument. Feedback was provided to the CS, and, as needed, another certification interview was scheduled. CSs were given up to three attempts to pass certification; however, all four CSs passed certification after their second interview.

Over the course of the MHSS, the veteran CSs also participated in refresher training to further enhance their skills and knowledge about the study. CS refresher trainings with Dr. First occurred in October 2008, September 2009, August 2010, and October 2011. The CS refresher trainings were 90- to 120-minute group conference calls that focused on clinical interviewing, the DSM-IV diagnostic criteria, and the GAF scale. In addition, the CSs all attended the MHSS CI training sessions provided by Dr. First.

6. Data Collection

This chapter describes the data collection procedures used on the Mental Health Surveillance Study (MHSS), including the administration, editing, and keying of the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002).

6.1 Schedule

The 2008, 2011, and 2012 MHSS samples were designed to yield 1,500 clinical interviews distributed across four calendar quarters with approximately 375 interviews occurring per quarter. The 2009 and 2010 MHSS samples were designed to yield 500 clinical interviews distributed across four calendar quarters with approximately 125 interviews occurring per quarter.

National Survey on Drug Use and Health (NSDUH) respondents were selected and recruited for the follow-up clinical interview at the end of the initial NSDUH interview. The clinical interviews were completed within 2 to 4 weeks following the completion of the NSDUH interview. For the first three quarters of each year, the probability of MHSS selection was maintained until the end of NSDUH data collection each quarter. Thus, MHSS data collection was scheduled to end 4 weeks beyond the NSDUH data collection end date each quarter. As shown in Table 6.1, for the fourth quarter of every year except 2012, the probability of selection for the clinical interview was set to zero in late November so that MHSS cases would not be sampled without adequate time for completion (by the main study data collection end date). In 2012, MHSS recruitment ended on the same day as NSDUH main study data collection, which was December 20, 2012. MHSS data collection was scheduled to end within 4 weeks of MHSS recruitment, which was by January 17, 2013.

Table 6.1 NSDUH and MHSS Data Collection End Date, by Year

Year	MHSS Recruitment End Date	NSDUH Data Collection End Date	MHSS Data Collection End Date
2008	November 21	December 22	December 19
2009	November 16	December 21	December 14
2010	November 29	December 20	December 20
2011	November 29	December 20	December 20
2012	December 20	December 20	January 17, 2013

MHSS = Mental Health Surveillance Study.

6.2 SCID Administration

As discussed in Section 4.1, NSDUH respondents were selected and recruited for the follow-up clinical interview at the end of the NSDUH main study. Respondents agreeing to participate in the follow-up interview were asked to provide contact information, including first name, telephone number, an alternate telephone number if available, and best days and times to call. The field interviewer (FI) entered the NSDUH respondent's contact information into the NSDUH laptop and transmitted the data to RTI International. The next step was for a data

collection manager (DCM) to review the MHSS case and assign it to a clinical interviewer (CI). The DCMs took into consideration the sample participant's time zone and best days and times for contact and the CI's availability and time zone when assigning cases. The CIs made the first attempt to contact MHSS cases within 24 hours of receiving the assigned case to schedule an appointment for the interview. If the case was not reached on the first call attempt, subsequent calls were made during the sample participant's preferred time frame. After CIs made several attempts without success during the time period specified during the NSDUH interview, calls were made at times outside of the preferred time frame in an attempt to reach the case.

At the scheduled interview time, the CI would attempt to contact the respondent by telephone to conduct the interview. As described in Section 4.2.1, CIs were instructed to verify that they were speaking to the respondent who had agreed to participate in the follow-up interview, obtain informed consent, and obtain permission to record the interview using the script provided (see Appendix C, Introduction pages 1-3). Permission to record the interview was not a requirement to participate in the study. CIs were instructed to include notes in the SCID about the respondent's reason for not giving permission to record, if known. To ensure confidentiality and privacy, no identifying information was written in the SCID or solicited during the recorded interview.

As discussed in Section 4.2.2, the follow-up interview was a modified version of the SCID (First, Spitzer, Gibbon, & Williams, 2002). The SCID is a semistructured diagnostic interview used to assess psychiatric disorders according to the criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994). As a semistructured clinical interview, the SCID contains structured, standardized questions that are read verbatim and sequentially, combined with unstructured follow-up questions that the CI tailors to the respondent based on clinical judgment and respondent answers. The SCID was administered over the telephone by CIs who had undergone extensive training with clinical supervisors (CSs) and the SCID's developer from Columbia University (see Chapter 5, Staff Training).

CIs were trained to use their clinical judgment to code each item based on the respondent's answers. Using the SCID booklet, each criterion symptom was coded as "1" (Absent or false), "2" (Subthreshold), "3" (Threshold or true), or "?" (Inadequate information). Each diagnosis assessed was coded as "1" (Absent) or "3" (Present). As discussed in Section 4.2.3, the CI would also score the Global Assessment of Functioning (GAF) on a scale of 1 (Persistent Danger to Self or Others) to 100 (Superior Functioning, No Symptoms). The GAF is based on the respondent's psychological, social, and occupational functioning. CIs scored the GAF based on the respondent's worst functioning in the past 12 months. At the conclusion of the SCID, the CI read the End of Interview Script (see Appendix C, End of Interview page K.4), which thanked the respondent for his or her time and mentioned the possibility of needing to speak with a counselor (i.e., "Sometimes the personal issues we've discussed cause people to become upset and in need of speaking with a counselor. If you are feeling upset or disturbed by the personal issues we have discussed in this interview and would like to speak with someone about your feelings, we suggest you call your doctor, counselor, or other treatment provider if you are currently under someone's care. If not, there is also a national lifeline number you can call. This number is on the receipt for the $30 you received for this interview from the interviewer who met with you earlier.") CIs ensured that the respondent had the toll-free number to the National Lifeline Network should he or she want to speak with a counselor and/or want

additional information about mental health services in his or her area. If the CI felt the respondent was at all in distress or potentially was a danger to himself or herself or others, a Distressed Respondent Protocol (DRP) was followed (Section 4.2.5).

Each clinical interview was completed over the telephone within 4 weeks of the date of the NSDUH interview. After the call ended, the CI completed the interviewer debriefing questions (see Appendix C, page X.1) within the SCID and reviewed the booklet to ensure that documentation was complete. As part of the interviewer debriefing section, CIs recorded information about problems encountered during the interview (i.e., distressed respondent, cognitive impairment, problems with privacy, comprehension, cooperation), stressors the respondents had experienced in the past 12 months, diagnosis that needed further assessment, and the overall validity of the SCID data. If CIs were uncertain of how to code a particular item or were feeling unsure about whether they had gathered enough information, CIs were instructed to consult with one of the CSs for his or her opinion within an hour of completing the interview. Within 48 hours after completion of the interview, the CI uploaded the recorded audio file to the Web-based Case Management System (CMS) and edited and shipped the paper SCID to RTI for final editing and keying in-house. Once shipped, the CI assigned the case a code of "80: Materials Shipment (SCID)" and entered the FedEx tracking number in the notes. The DCM then tracked the shipment to ensure the SCID arrived at RTI.

6.2.1 SCID Length

After obtaining informed consent, the CIs recorded the interview start time and AM/PM designation at the beginning of the Overview section on the SCID (see Appendix C, Overview page i). After reading the End of Interview script at the conclusion of the SCID, the CIs recorded the interview end time and AM/PM designation in the End of Interview section on the SCID (see Appendix C, page K.4). The interview start and end times recorded by the CIs on the SCID were keyed and included in the data file.

As shown in Table 6.2, the average time to complete the SCID during the 2008-2012 MHSS was approximately 72 minutes. The median was 60 minutes. Completed interviews in which a breakoff occurred (n = 275) and/or completed interviews with incomplete timing data (n = 55), such as the missing start or end time or the AM/PM designation, were excluded from the timing analysis and summary statistics in Table 6.2.

Breakoff interviews occurred when the CI started the interview, but the respondent was unable to complete the interview at that time (e.g., because of interview length, personal nature of the questions, feeling distressed, cognitive impairment) or the phone call was disconnected. If the respondent requested to end the interview, the CI was instructed to inquire about the respondent's concerns, address problems if possible, and try to schedule a convenient time for the respondent to complete the interview as soon as possible but still within 4 weeks of completing the NSDUH interview. If the respondent's concerns could not be satisfactorily addressed, the interview was terminated, and no further contact was made. More information is provided on the breakoff protocol for cognitive impairment and distressed respondent situations in Sections 4.2.4 and 4.2.5. In the cases in which the phone connection was lost, the CI made multiple attempts to reconnect with the respondent the same day and future days within 4 weeks after completing the NSDUH interview. Of the 507 breakoff interviews on the 2008-2012 MHSS, 232 breakoff interviews were not completed and assigned a final status code of “73: Breakoff, Partial

Interview," and 275 breakoff interviews were resumed and completed later. In these situations, the breakoff interview time and resumed interview time were not captured in the SCID; thus, breakoff interviews were excluded from the timing data in Table 6.2.

Table 6.2 2008-2012 NSDUH MHSS SCID Length

	2008	2009	2010	2011	2012	Total (2008-2012)
Sample Used in Analysis	1,406	482	490	1,399	1,546	5,323
Excluded Records*	94	38	26	96	76	330
Summary Statistics (Minutes)						
Mean	68.95	72.01	73.82	75.26	71.84	72.18
Variance	3,384.24	5,496.82	5,237.02	7,040.61	4,099.42	4,916.49
Standard Deviation	58.17	74.14	72.37	83.91	64.03	70.12
Quartiles						
Maximum	835.00	877.00	844.00	816.00	834.00	877.00
Quartile 3	83.00	82.00	87.00	85.00	85.00	84.00
Median	60.00	59.00	61.00	60.00	62.00	60.00
Quartile 1	43.00	44.00	45.00	43.00	44.00	44.00
Minimum	1.00	22.00	16.00	8.00	19.00	1.00
Mode	60.00	45.00	55.00	60.00	40.00	60.00
Range	834.00	855.00	828.00	808.00	815.00	876.00
Percentiles						
99%	169.00	210.00	190.00	750.00	174.00	195.00
95%	124.00	130.00	128.00	135.00	124.00	128.00
90%	107.00	110.00	112.00	110.00	109.00	109.00
10%	33.00	35.00	33.00	33.00	34.00	33.00
5%	29.00	30.00	29.00	29.00	30.00	29.00
1%	22.00	25.00	23.00	22.00	24.00	23.00
Extremes						
5 Highest (Highest)	835.00	877.00	844.00	816.00	834.00	877.00
	798.00	830.00	800.00	808.00	803.00	844.00
	788.00	770.00	788.00	800.00	787.00	835.00
	770.00	756.00	751.00	794.00	770.00	834.00
	760.00	210.00	190.00	793.00	767.00	830.00
5 Lowest	19.00	25.00	23.00	18.00	21.00	16.00
	19.00	24.00	22.00	17.00	21.00	12.00
	19.00	24.00	21.00	12.00	20.00	9.00
	17.00	23.00	21.00	9.00	20.00	8.00
(Lowest)	1.00	22.00	16.00	8.00	19.00	1.00

MHSS = Mental Health Surveillance Study; SCID = Structured Clinical Interview for DSM-IV.

* Completed interviews in which a breakoff occurred and/or completed interviews with incomplete timing data were excluded from the timing analysis and summary statistics. Completed interviews with complete, but extreme, timing data were included in the analysis.

6.3 SCID Data Processing

When the SCID was received at RTI, it was assigned a status code of "81: SCID Received at Regent." The SCID was then reviewed by CSs, technical editors, and keying clerks, following the editing and keying process described below.

6.3.1 Clinical Editing

When the SCID was delivered to the MHSS CSs, it was assigned a status code of "82: SCID Received at Editing." The CSs reviewed all of the data collected in every SCID booklet, item by item, comparing the notes provided by the CI with the diagnostic rating and listening to parts or all of the accompanying audio files as needed to ensure confidence in the data. The audio file (if recorded) was matched with the corresponding paper-and-pencil SCID booklet. For full reviews, both the audio recordings and the SCID booklet were reviewed in their entirety. A newly hired CI received full reviews of their first two interviews or until he or she demonstrated proficiency in administering all SCID modules. Once a CI was deemed proficient, 10 percent of the remaining interviews for each CI were randomly selected for a full review of the audio recordings and SCID booklet in their entirety.

If the interview was selected for full review, the audio file(s) were listened to in their entirety and compared against the CI's notes. If the interview was not selected for a full review, then the CSs completed a partial review of the audio file(s) to include the SCID's overview section and any other portions of the audio file(s) that needed review to clarify the responses, to supplement the notes, or to support the CI's ratings in the SCID booklet. Cases that presented a more complex clinical picture generally warranted full reviews as well.

As part of the review, CSs ensured that CIs followed project protocol for administering the interview and provided accurate and sufficient notes in the SCID booklet. The CSs also evaluated the CIs' clinical interviewing techniques, diagnostic skills, and the extent to which they captured the overall clinical picture relative to the symptoms, diagnoses, and GAF score. In the back of the SCID booklet, the CS rated the overall validity of the data and noted if there were any other disorders that needed further assessment, such as a disorder that was not assessed in the study or a disorder for which more information was needed for diagnosis. For each interview, the CSs completed a clinical editing form, noting strengths and areas for improvement for the CI (Appendix G). Completed SCID editing forms were filed in separate folders for each CI in a locked filing cabinet.

Completed interviews that one CS considered to be of questionable validity were subject to an independent second review by at least one other CS. To address the potential bias to exclude more complex or unusual cases and those with a difficult respondent, the level of difficulty and overall confidence in the case were considered before making a final disposition. If all CSs agreed, then the case was removed from the study. In these situations, a code of "90: SCID Not Keyed - Reason Unspecified" (before 2010) or "90.X" (after 2010; "X" depending on the reason) was entered into the CMS. Any relevant feedback was given directly to the CI.

Once clinical editing was completed, the corresponding code of "88: SCID Delivered to Technical Editing" was entered into the CMS, and the SCID booklet was given to the technical editors. Refer to Chapter 8 for more information on data collection quality control.

6.3.2 Technical Editing

Once clinical editing was completed, the technical editors reviewed each SCID to check for missing or inconsistent responses (i.e., compare variables that should match because their values are dependent on the value of another variable) and to verify that conditions stated in the

"Skip Logic" criteria were true (i.e., to check that the CI followed the correct skip patterns when the response to one question directed the CI to the next question, a follow-up question, or a question in another module) (see Appendix H, SCID Technical Editing Guide).

As part of this technical review, technical editors checked that (1) SCID responses were recorded correctly on the front cover of the SCID booklet (e.g., FI ID, Quest ID, date of interview), (2) all required variables were recorded and within the range for the response, (3) fields were consistent with screening variables, (4) skip patterns were followed correctly, and (5) conditions were met for a code "3" (present) for symptom criteria and diagnostic variables. For example, the SCID screening questions, which were asked at the end of the SCID overview, are coded on the last two pages of the overview module and at the beginning of the respective diagnostic modules. If the screening question for a specific disorder is coded as "1" (absent) in the overview, then it must also be coded as "1" in the section that assesses that particular disorder. Another example would be Module A and Module D, both of which are used for coding mood disorders at the level of symptoms (Module A) and diagnoses (Module D). The values for the variables in Module D (Mood Differential) are contingent upon the values in Module A (Mood Episodes). If the criteria for a specific mood episode in Module A are met and therefore coded as "3" (present), then the variable for coding that same mood episode in Module D would also be coded as "3" (present).

Any inconsistencies or cases in which the conditions were not true were noted on a technical editing form (see Appendix H, SCID Failed Edit Worksheet). The technical editing form and its respective SCID were returned to the CSs for final review and correction. Once the technical editors had finished, a code of "89: Technical Editing Complete" was recorded in the CMS.

6.3.3 Keying

Once technical editing was completed and inconsistencies were resolved, the data clerks keyed the SCID data into a computer database and stored the keyed data in an electronic file. All codes that did not meet the keying specification criteria (for consistency and skip logic conditions) were flagged by a machine editing program. A report of flagged cases was sent to the CSs indicating the line(s) where the errors were suspected. SCID booklets flagged with potential problems were retrieved from storage and compared to the computer-generated report. Those line(s) were reviewed again to ensure accuracy in both clinical and technical editing. Any errors were communicated to the dataset manager to be corrected in the electronic file if needed.

6.3.4 Storage and Destruction

All hard copy SCIDs, audio files, and supporting documentation were securely stored following data collection. With the exception of the 2008 SCIDs, which were retained for 6 months following the end of data collection, the SCIDs were retained for 18 months after the end of the data collection period. The audio files for the 2008-2010 Quarter 2 SCID interviews were retained for 6 months after the end of the data collection period. Beginning in Quarter 3 of 2010, Institutional Review Board (IRB) approval was received to retain the audio files for 18 months after the end of the data collection period. The Introduction to Clinical Follow-Up Interview script was modified beginning with interviews conducted in Quarter 3 of 2010 to reflect deleting the audio files within 18 months after the end of the data collection period. This

extension allowed for more time in the data processing schedule for coding and keying additional SCID variables. The SCIDs were kept in a secured storage facility until destroyed. All materials set for destruction were directly sent to a secured shredding facility.

6.4 Data Collection Management and Reporting

Management of MHSS data collection placed a strong emphasis on production, cost-effectiveness, data quality, and communication. To stay informed of data collection status, project management maintained a series of routine meetings:

- DCMs conducted a weekly conference call with every CI with active case work to gain feedback and discuss production, methods to maximize efficiency (e.g., while administering the SCID, listen closely, conduct proper probing, provide adequate documentation, avoid asking vague and/or inappropriate questions), and administrative issues.
- CSs conducted conference calls continuously throughout the quarter with every CI with active case work to review and provide specific feedback on data quality, distressed respondent issues, and proper interviewing techniques.
- The mental health task leader conducted weekly conference calls with the CSs to review data quality, distressed respondent issues, editing and keying completion rates, and interview quality.
- The operations task leader conducted monthly meetings with the CSs, DCMs, sampling staff, training and materials development staff, and other key MHSS project staff to discuss the schedule and upcoming operational activities and deadlines.
- MHSS task leaders met with Substance Abuse and Mental Health Services Administration staff every 2 weeks to provide a status update on operational activities.

In addition to these formal meetings, staff communicated frequently through e-mail and by telephone. Project management, DCMs, and CSs monitored the status of data collection through daily reports provided on the NSDUH CMS. The MHSS CMS included the following functions and reports:

- case status—to view cases by status code and date;
- record of calls (ROC)—to view contact attempts and outcomes for each case;
- case transfer/assignment—to assign and transfer cases to CIs;
- clinical interview file repository—for CSs to listen to recorded interviews;
- incident reports—for project management and CSs to view incident reports completed for distressed respondent cases;
- electronic timesheet—to review and approve timesheets completed by CSs and CIs; and
- monitoring reports, including the Executive Summary Report (with production and cost data), Recruitment Report (a list of hired CIs), Pending Clinical Interview Report (with the status code and date, the number of ROCs, and a time window to complete

each interview), MHSS Sample Monitoring Report (Section 6.4.1), and Clinical Interview Current Status Report (Section 6.4.2).

6.4.1 MHSS Sample Monitoring Report

The daily MHSS Sample Monitoring Report documented the MHSS eligibility, selection, agreement, and completion rates by the Kessler-6 score. This report was posted daily to the CMS and was accessible to supervisory and management staff. The MHSS Sample Monitoring Report enabled RTI statisticians to monitor all key variables associated with the MHSS sampling yields and take action as needed to ensure the project reached data collection goals every quarter. The report also enabled management staff to anticipate production and CI staffing needs. Table 6.3 provides an example of the MHSS Sample Monitoring Report from the 2012 MHSS, Quarter 4.

Table 6.3 MHSS Sample Data: 2012 Quarter 4

Kessler-6 Score	Interview Respondents (Aged 18+)	Eligible for MHSS	Selected for Clinical Follow-up		Agree to Clinical Follow-up		Completed Clinical Follow-up	
			Count	%	Count	%	Count	%
0-3	5,035	4,786	216	4.5	158	73.1	129	81.6
4-5	1,488	1,451	59	4.1	50	84.7	46	92.0
6-7	1,068	1,045	35	3.3	28	80.0	23	82.1
8-9	714	696	42	6.0	37	88.1	34	91.9
10-11	585	580	47	8.1	37	78.7	28	75.7
12-15	918	907	114	12.6	107	93.9	86	80.4
16+	1,010	995	137	13.8	126	92.0	86	68.3
Total	10,818	10,460	650	6.2	543	83.5	432	79.6

MHSS = Mental Health Surveillance Study.

6.4.2 Clinical Interview Current Status Report

The Clinical Interview Current Status Report provided the current status of each MHSS case and a summary of MHSS interview completion counts by quarter. By utilizing the ROC system within the CMS, this status report tracked the progress of every case, from when it was first activated and transmitted to RTI to the final receipt, editing, and keying of completed SCIDs. This report enabled management staff to clearly monitor daily contacts made by CIs and to monitor the multistep editing and keying process of incoming SCIDs for completeness and accuracy. At the conclusion of each quarter, this report served as a final progress check to ensure every case was processed and assigned a final code. Appendix I provides an example of the Clinical Interview Current Status Report from the 2012 MHSS.

6.5 Problems Encountered

For data collection, the CIs encountered problems with the audio recording of SCID interviews over the telephone. The audio recording system implemented in 2008 required the CIs to use a recording control adaptor between a wall jack and telephone if an analog line or between the telephone handset and telephone if a digital line. Using this method, the CI's voice was clearly recorded, but the respondent's voice was difficult to hear at times. As discussed in Section 4.3.2, after testing, the CIs began using the Internet-based phone provider Skype in June 2011 for the interview contact with respondents. The quality of the recordings was significantly improved.

The CIs also encountered problems with the DRP. As described in Chapter 7, there were times when the CIs did not fully adhere to the IRB-approved protocol for handling distressed respondents. Violations of the DRP occurred when the CI did not recognize the respondent's signs of emotional distress (e.g., sadness, irritability, agitation) or if the CI did not fully assess the respondent's suicidal ideation. As discussed in Chapter 4, DRP violations also occurred due to unforeseen logistical problems with the emergency contact protocol. The DRP was updated to address these issues. The DRP was reviewed and practiced during all veteran and refresher CI training sessions, with special attention on how to listen for and adequately assess suicidal or homicidal thoughts and how to recognize and address emotional distress at any point of the interview (see Chapter 5, Staff Training). Whenever protocol violations occurred, the CI was immediately contacted and either retrained or terminated.

7. Data Collection Results

This chapter describes the results from the Mental Health Surveillance Study (MHSS) data collection from 2008 to 2012.

7.1 Response Rates

The 2008, 2011, and 2012 MHSS samples were designed to yield 1,500 clinical follow-up interviews distributed across four calendar quarters with approximately 375[5] interviews occurring per quarter. The 2009 and 2010 MHSS samples were designed to yield 500 clinical follow-up interviews distributed across four calendar quarters with approximately 125 interviews occurring per quarter. As discussed in Section 6.1, for every year except 2012, the probability of selection of the National Survey on Drug Use and Health (NSDUH) interview respondents for the clinical follow-up survey in Quarter 4 was set to zero in late November. Respondents who received a zero probability of selection but would have been selected based on their Kessler-6 score, World Health Organization Disability Assessment Schedule score, and age group were referred to as the zero probability cases.

Table 7.1 provides a summary of the MHSS response rates by year. A total of 220,219 NSDUH interview respondents were eligible for the MHSS follow-up interview from 2008 to 2012. Eligible NSDUH respondents were aged 18 years or older and completed the main study interview in English. Over the 5 years of the MHSS, 8,629 respondents were selected for the clinical follow-up interview. Of those selected, 7,222 agreed to participate, for an agreement rate of 83.69 percent.[6] Of the respondents who agreed to participate, a total of 5,653 (78.27 percent) MHSS valid interviews were completed during the entire MHSS data collection period.

In 2008, a total of 2,331 NSDUH respondents were sampled, 47 of whom were zero probability cases and treated as nonrespondents. A total of 1,973 selected respondents agreed to participate, for an agreement rate of 84.6 percent. Excluding the zero probability cases, 2,284 respondents were selected, and 1,973 (86.3 percent) agreed to participate. A total of 1,500 (76.0 percent) of those respondents who agreed to participate completed a valid clinical interview. The overall completion rate was 64.4 percent and does not incorporate the NSDUH main study nonresponse rates. A summary of the 2008 MHSS respondents by quarter is included in Table 7.2.

In 2009, a total of 789 NSDUH respondents were sampled, 21 of whom were zero probability cases and treated as nonrespondents. A total of 665 selected respondents agreed to participate, for an agreement rate of 84.3 percent. Excluding the zero probability cases, 768 respondents were selected, and 665 (86.6 percent) agreed to participate. A total of 520 (78.2 percent) of the respondents who agreed to participate completed a valid clinical interview. The overall completion rate was 65.9 percent and does not incorporate the NSDUH main study nonresponse rates. A summary of the 2009 MHSS respondents by quarter is included in Table 7.3.

---

[5] The goals for Quarters 3 and 4 of 2012 were increased from 375 to 385 each quarter.

[6] NSDUH respondents who agreed to clinical follow-up at the time of their main study interview are classified as agreeing to participate.

Table 7.1 2008-2012 MHSS Response Rate Summary

Design Parameter	2008	2009	2010	2011	2012
Interview Respondents Aged 18 or Older	45,678	45,609	45,844	46,599	45,836
Eligible for MHSS[1]	43,593	43,708	43,958	44,740	44,220
Eligibility Rate	*0.9544*	*0.9583*	*0.9589*	*0.9601*	*0.9647*
Selected for Telephone Clinical Follow-up[2]	2,331	789	768	2,277	2,464
Zero Probability Cases	47	21	4	41	0
Agreed to Clinical Follow-up	1,973	665	640	1,881	2,063
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8464*	*0.8428*	*0.8333*	*0.8261*	*0.8373*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8638*	*0.8659*	*0.8377*	*0.8412*	*0.8373*
Completed Clinical Interviews	1,500	520	516	1,495	1,622
Clinical Interview Completion Rate	*0.7603*	*0.7820*	*0.8063*	*0.7948*	*0.7862*
Overall Clinical Follow-up Response Rate[3]	*0.6435*	*0.6591*	*0.6719*	*0.6566*	*0.6583*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] Includes zero probability cases and treats them as nonrespondents.

Table 7.2 2008 MHSS, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,692	12,816	11,355	10,815	45,678
Eligible for MHSS[1]	10,215	12,148	10,849	10,381	43,593
Eligibility Rate	*0.9554*	*0.9479*	*0.9554*	*0.9599*	*0.9544*
Selected for Telephone Clinical Follow-up[2]	696	529	485	621	2,331
Zero Probability Cases	0	0	0	47	47
Agreed to Clinical Follow-up	586	462	416	509	1,973
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8420*	*0.8733*	*0.8577*	*0.8196*	*0.8464*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8420*	*0.8733*	*0.8577*	*0.8868*	*0.8638*
Completed Clinical Interviews	467	361	317	355	1,500
Clinical Interview Completion Rate	*0.7969*	*0.7814*	*0.7620*	*0.6974*	*0.7603*
Overall Clinical Follow-up Response Rate[3]	*0.6710*	*0.6824*	*0.6536*	*0.5717*	*0.6435*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] Includes zero probability cases and treats them as nonrespondents.

In 2010, a total of 768 NSDUH respondents were sampled, 4 of whom were zero probability cases and treated as nonrespondents. A total of 640 selected respondents agreed to participate, for an agreement rate of 83.3 percent. Excluding the zero probability cases, 764 respondents were selected, and 640 (83.8 percent) agreed to participate. A total of 516 (80.6 percent) of the respondents who agreed to participate completed a valid clinical interview. The overall completion rate was 67.2 percent and does not incorporate the NSDUH main study nonresponse rates. A summary of the 2010 MHSS respondents by quarter is included in Table 7.4.

Table 7.3 2009 MHSS, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	11,403	11,963	11,264	10,979	45,609
Eligible for MHSS[1]	10,930	11,452	10,786	10,540	43,708
Eligibility Rate	*0.9585*	*0.9573*	*0.9576*	*0.9600*	*0.9583*
Selected for Telephone Clinical Follow-up[2]	182	192	211	204	789
Zero Probability Cases	0	0	0	21	21
Agreed to Clinical Follow-up	156	167	183	159	665
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8571*	*0.8698*	*0.8673*	*0.7794*	*0.8428*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8571*	*0.8698*	*0.8673*	*0.8689*	*0.8659*
Completed Clinical Interviews	123	125	142	130	520
Clinical Interview Completion Rate	*0.7885*	*0.7485*	*0.7760*	*0.8176*	*0.7820*
Overall Clinical Follow-up Response Rate[3]	*0.6758*	*0.6510*	*0.6730*	*0.6373*	*0.6591*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] Includes zero probability cases and treats them as nonrespondents.

Table 7.4 2010 MHSS, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,877	12,102	11,844	11,021	45,844
Eligible for MHSS[1]	10,446	11,608	11,341	10,563	43,958
Eligibility Rate	*0.9604*	*0.9592*	*0.9575*	*0.9584*	*0.9589*
Selected for Telephone Clinical Follow-up[2]	190	246	175	157	768
Zero Probability Cases	0	0	0	4	4
Agreed to Clinical Follow-up	163	198	146	133	640
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8579*	*0.8049*	*0.8343*	*0.8471*	*0.8333*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8579*	*0.8049*	*0.8343*	*0.8693*	*0.8377*
Completed Clinical Interviews	132	157	115	112	516
Clinical Interview Completion Rate	*0.8098*	*0.7929*	*0.7877*	*0.8421*	*0.8063*
Overall Clinical Follow-up Response Rate[3]	*0.6947*	*0.6382*	*0.6571*	*0.7134*	*0.6719*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] Includes zero probability cases and treats them as nonrespondents.

In 2011, a total of 2,277 NSDUH respondents were sampled, 41 of whom were zero probability cases and treated as nonrespondents. A total of 1,881 selected respondents agreed to participate, for an agreement rate of 82.6 percent. Excluding the zero probability cases, 2,236 respondents were selected, and 1,881 (84.1 percent) agreed to participate. A total of 1,495 (79.5 percent) of the respondents who agreed to participate completed a valid clinical interview. The overall completion rate was 65.7 percent and does not incorporate the NSDUH main study nonresponse rates. A summary of the 2011 MHSS respondents by quarter is included in Table 7.5.

Table 7.5 2011 MHSS, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,840	12,481	12,170	11,108	46,599
Eligible for MHSS[1]	10,392	11,974	11,665	10,709	44,740
Eligibility Rate	*0.9587*	*0.9594*	*0.9585*	*0.9641*	*0.9601*
Selected for Telephone Clinical Follow-up[2]	543	672	531	531	2,277
Zero Probability Cases[3]	15	0	0	26	41
Agreed to Clinical Follow-up	450	561	449	421	1,881
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8287*	*0.8348*	*0.8456*	*0.7928*	*0.8261*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8523*	*0.8348*	*0.8456*	*0.8337*	*0.8412*
Completed Clinical Interviews	363	436	359	337	1,495
Clinical Interview Completion Rate	*0.8067*	*0.7772*	*0.7996*	*0.8005*	*0.7948*
Overall Clinical Follow-up Response Rate[4]	*0.6685*	*0.6488*	*0.6761*	*0.6347*	*0.6566*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] At the beginning of Quarter 1, 15 interview respondents who should have been selected for the MHSS were inadvertently assigned a zero probability of selection.

[4] Includes zero probability cases and treats them as nonrespondents.

In 2012, a total of 2,464 NSDUH respondents were sampled. No cases were marked as zero probability cases in 2012. A total of 2,063 selected respondents agreed to participate, for an agreement rate of 83.7 percent. A total of 1,622 (78.6 percent) of the respondents who agreed to participate completed a valid clinical interview. The overall completion rate was 65.8 percent and does not incorporate the NSDUH main study nonresponse rates. A summary of the 2012 MHSS respondents by quarter is included in Table 7.6.

Table 7.6 2012 MHSS, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,894	12,144	11,997	10,801	45,836
Eligible for MHSS[1]	10,488	11,699	11,591	10,442	44,220
Eligibility Rate	*0.9627*	*0.9634*	*0.9662*	*0.9668*	*0.9647*
Selected for Telephone Clinical Follow-up[2]	585	596	634	649	2,464
Zero Probability Cases[3]	0	0	0	0	0
Agreed to Clinical Follow-up	491	497	532	543	2,063
Percentage Agreeing to Clinical Follow-up (including zero probability cases)	*0.8393*	*0.8339*	*0.8391*	*0.8367*	*0.8373*
Percentage Agreeing to Clinical Follow-up (excluding zero probability cases)	*0.8393*	*0.8339*	*0.8391*	*0.8367*	*0.8373*
Completed Clinical Interviews	384	379	434	425	1,622
Clinical Interview Completion Rate	*0.7821*	*0.7626*	*0.8158*	*0.7827*	*0.7862*
Overall Clinical Follow-up Response Rate	*0.6564*	*0.6359*	*0.6845*	*0.6549*	*0.6583*

MHSS = Mental Health Surveillance Study.

[1] Respondents aged 18 or older who completed their main study interview in English are eligible to be selected for the MHSS.

[2] Includes cases assigned a zero probability of selection that would have been selected based on their Kessler-6 and World Health Organization Disability Assessment Schedule scores and age groups.

[3] There were no zero probability cases in 2012.

7.2 Number of Contacts

After a respondent was selected at the end of the NSDUH interview to complete a clinical interview, the case was transmitted back to RTI International and assigned to a clinical interviewer (CI). The CIs made the first attempt to contact a case within 24 hours of receiving the assigned case to schedule an appointment for the interview. If the case was not reached on the first call attempt, subsequent calls were made during the case's preferred time frame provided during the NSDUH interview. If a CI made several attempts without success during the time period specified for the case, calls were made at times outside of the case's preferred time frame in an attempt to reach the case. The data collection managers (DCMs) completed reviews of the record of calls to ensure that call attempts were being made at appropriate times and provided guidance on the best days and times to recontact cases. An MHSS case was considered final once a respondent completed the clinical interview or the case was finalized as a nonrespondent (i.e., cases who refused or could not be contacted and therefore did not complete the interview).

Table 7.7 provides the number of call attempts made to contact the MHSS cases across all survey years. In all survey years, about 25 percent of interviews were completed in one or two call attempts, more than 56 percent were completed within four call attempts, and 75 percent within six to seven call attempts. Each year, the CIs made 10 or more call attempts to complete 11 to 14 percent of the clinical interviews, with more than 95 percent of interviews completed within 15 call attempts. The number of call attempts made to contact the MHSS case remained about the same for each year of the MHSS.

Table 7.8 provides the number of call attempts made to contact the MHSS cases finalized as nonrespondents across all survey years. In 2008, the CIs determined 65 percent of these cases to be nonrespondents within 15 contact attempts. The percentage of nonrespondents finalized within 15 contact attempts declined each year of the MHSS. This means that CIs made more call attempts to cases each year before determining the case to be a nonrespondent. In 2012, only 32 percent of nonrespondent cases were finalized within 15 call attempts, and 59 percent of cases were classified as nonrespondents within 20 call attempts.

Table 7.7 Number of Call Attempts Made To Contact MHSS Respondents: 2008-2012

Call Attempts (#)	2008			2009			2010			2011			2012		
	Total Respondents (#)	%	Cumula-tive %	Total Respondents (#)	%	Cumula-tive %	Total Respondents (#)	%	Cumula-tive %	Total Respondents (#)	%	Cumula-tive %	Total Respondents (#)	%	Cumula-tive %
1	65	4.27	4.27	6	1.14	1.14	13	2.50	2.50	66	4.36	4.36	64	3.89	3.89
2	308	20.22	24.49	114	21.59	22.73	111	21.39	23.89	335	22.13	26.49	351	21.32	25.21
3	297	19.50	43.99	110	20.83	43.56	101	19.46	43.35	288	19.02	45.51	333	20.23	45.44
4	197	12.93	56.93	75	14.20	57.77	73	14.07	57.42	220	14.53	60.04	217	13.18	58.63
5	151	9.91	66.84	62	11.74	69.51	43	8.29	65.70	136	8.98	69.02	151	9.17	67.80
6	101	6.63	73.47	32	6.06	75.57	43	8.29	73.99	85	5.61	74.64	108	6.56	74.36
7	78	5.12	78.59	24	4.55	80.11	29	5.59	79.58	84	5.55	80.13	89	5.41	79.77
8	71	4.66	83.26	26	4.92	85.04	17	3.28	82.85	49	3.24	83.42	62	3.77	83.54
9	60	3.94	87.20	20	3.79	88.83	17	3.28	86.13	51	3.37	86.79	55	3.34	86.88
10-15	145	9.52	96.72	48	9.09	97.92	49	9.44	95.57	134	8.85	95.64	154	9.36	96.23
16-20	33	2.17	98.88	10	1.89	99.81	18	3.47	99.04	48	3.17	98.81	39	2.37	98.60
21-25	10	0.66	99.54	1	0.19	100.00	3	0.58	99.61	11	0.73	99.54	8	0.49	99.09
26-30	5	0.33	99.87	0	0.00	100.00	2	0.39	100.00	7	0.46	100.00	12	0.73	99.82
31+	2	0.13	100.00	0	0.00	100.00	0	0.00	100.00	0	0.00	100.00	3	0.18	100.00
Total	1,523	100.00		528	100.00		519	100.00		1,514	100.00		1,646	100.00	

MHSS = Mental Health Surveillance Study.

Table 7.8 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2008-2012

Call Attempts (#)	2008			2009			2010			2011			2012		
	Total Non-respondents (#)	%	Cumula-tive %	Total Non-respondents (#)	%	Cumula-tive %	Total Non-respondents (#)	%	Cumula-tive %	Total Non-respondents (#)	%	Cumula-tive %	Total Non-respondents (#)	%	Cumula-tive %
1-5	42	9.15	9.15	17	12.41	12.41	11	8.94	8.94	27	7.09	7.09	38	9.00	9.00
6-10	115	25.05	34.20	19	13.87	26.28	8	6.5	15.45	41	10.76	17.85	33	7.82	16.82
11-15	145	31.59	65.80	40	29.20	55.47	33	26.83	42.28	82	21.52	39.37	64	15.17	31.99
16-20	61	13.29	79.08	40	29.20	84.67	25	20.33	62.60	102	26.77	66.14	117	27.73	59.72
21-25	47	10.24	89.32	18	13.14	97.81	32	26.02	88.62	82	21.52	87.66	101	23.93	83.65
26-30	23	5.01	94.34	2	1.46	99.27	13	10.57	99.19	37	9.71	97.38	50	11.85	95.50
31+	26	5.66	100.00	1	0.73	100.00	1	0.81	100.00	10	2.62	100.00	19	4.50	100.00
Total	459	100.00		137	100.00		123	100.00		381	100.00		422	100.00	

MHSS = Mental Health Surveillance Study.

Table 7.9 shows the number of call attempts made to contact the MHSS respondents in 2008. Almost 25 percent of the clinical interviews were completed in one or two call attempts. The CIs made 10 or more call attempts to complete 13 percent of the clinical interviews, and more than 96 percent of interviews were completed within 15 call attempts.

Table 7.10 shows the number of call attempts made to contact the MHSS cases finalized as nonrespondents in 2008. For the vast majority of these cases, the CIs were unable to make contact with the case. The CIs determined 79 percent of these cases to be nonrespondents within 20 contact attempts, and 89 percent of interviews were classified as nonrespondents within 25 contacts.

Table 7.9 Number of Call Attempts Made To Contact MHSS Respondents: 2008

Call Attempts (#)	Total Respondents (#)	Total Respondents (%)	Cumulative Total Respondents (%)	Code 70 (Interview Complete, Audio Recorded)	Code 71 (Interview Complete, No Audio—Refusal)	Code 72 (Interview Complete, No Audio—Technical Problem)
1	65	4.27	4.27	63	1	1
2	308	20.22	24.49	283	16	9
3	297	19.50	43.99	277	7	13
4	197	12.93	56.93	175	13	9
5	151	9.91	66.84	137	6	8
6	101	6.63	73.47	93	4	4
7	78	5.12	78.59	70	5	3
8	71	4.66	83.26	63	5	3
9	60	3.94	87.20	53	7	0
10-15	145	9.52	96.72	132	6	7
16-20	33	2.17	98.88	29	2	2
21-25	10	0.66	99.54	9	1	0
26-30	5	0.33	99.87	4	0	1
31+	2	0.13	100.00	2	0	0
Total	1,523	100.00		1,390	73	60

MHSS = Mental Health Surveillance Study.

Table 7.10 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2008

Call Attempts (#)	Total Non-respondents (#)	Total Non-respondents (%)	Cumulative Total Non-respondents (%)	Code 73 (Breakoff, Partial Interview)	Code 74 (Unable to Contact)	Code 75 (Phone Number Problem)	Code 76 (Refusal)	Code 77 (Other)
1-5	42	9.15	9.15	12	3	16	4	7
6-10	115	25.05	34.20	20	61	27	4	3
11-15	145	31.59	65.80	9	119	13	1	3
16-20	61	13.29	79.08	11	41	6	0	3
21-25	47	10.24	89.32	2	39	4	1	1
26-30	23	5.01	94.34	3	17	3	0	0
31+	26	5.66	100.00	3	23	0	0	0
Total	459	100.00		60	303	69	10	17

MHSS = Mental Health Surveillance Study.

Table 7.11 shows the number of call attempts made to contact the MHSS respondents in 2009. The CIs completed almost 23 percent of the clinical interviews in one or two call attempts. The CIs made 10 or more call attempts to complete 11 percent of the clinical interviews, and more than 97 percent of cases were completed within 15 call attempts.

Table 7.11 Number of Call Attempts Made To Contact MHSS Respondents: 2009

Call Attempts (#)	Total Respondents (#)	Total Respondents (%)	Cumulative Total Respondents (%)	Code 70 (Interview Complete, Audio Recorded)	Code 71 (Interview Complete, No Audio—Refusal)	Code 72 (Interview Complete, No Audio—Technical Problem)
1	6	1.14	1.14	6	0	0
2	114	21.59	22.73	109	5	0
3	110	20.83	43.56	108	2	0
4	75	14.20	57.77	72	3	0
5	62	11.74	69.51	57	5	0
6	32	6.06	75.57	32	0	0
7	24	4.55	80.11	23	1	0
8	26	4.92	85.04	25	1	0
9	20	3.79	88.83	17	3	0
10-15	134	9.09	97.92	39	7	2
16-20	48	1.89	99.81	9	1	0
21-25	10	0.19	100.00	1	0	0
26-30	1	0.00	100.00	0	0	0
31+	0	0.00	100.00	0	0	0
Total	528	100.00		498	28	2

MHSS = Mental Health Surveillance Study.

Table 7.12 shows the number of call attempts made to contact the MHSS cases finalized as nonrespondents in 2009. For the majority of these cases, the CIs were unable to make contact with the case. The CIs determined 84 percent of these cases to be nonrespondents within 20 contact attempts, and almost 98 percent of interviews were classified as nonrespondents within 25 contacts.

Table 7.12 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2009

Call Attempts (#)	Total Non-respondents (#)	Total Non-respondents (%)	Cumulative Total Non-respondents (%)	Code 73 (Breakoff, Partial Interview)	Code 74 (Unable to Contact)	Code 75 (Phone Number Problem)	Code 76 (Refusal)	Code 77 (Other)
1-5	17	12.41	12.41	4	1	8	1	3
6-10	19	13.87	26.28	3	7	7	0	2
11-15	40	29.20	55.47	9	25	3	3	0
16-20	40	29.20	84.67	4	33	3	0	0
21-25	18	13.14	97.81	1	17	0	0	0
26-30	2	1.46	99.27	0	2	0	0	0
31+	1	0.73	100.00	0	1	0	0	0
Total	137	100.00		21	86	21	4	5

MHSS = Mental Health Surveillance Study.

Table 7.13 shows the number of call attempts made to contact the MHSS respondents in 2010. Almost 24 percent of the clinical interviews were completed in one or two call attempts. The CIs made 10 or more call attempts to complete 14 percent of the clinical interviews, and more than 95 percent of cases were completed within 15 call attempts.

Table 7.14 shows the number of call attempts made to contact the MHSS cases finalized as nonrespondents in 2010. For the majority of these cases, the CIs were unable to make contact with the case. The CIs determined 62 percent of these cases to be nonrespondents within 20 contact attempts, and 88 percent of interviews were classified as nonrespondents within 25 contacts.

Table 7.13 Number of Call Attempts Made To Contact MHSS Respondents: 2010

Call Attempts (#)	Total Respondents (#)	Total Respondents (%)	Cumulative Total Respondents (%)	Code 70 (Interview Complete, Audio Recorded)	Code 71 (Interview Complete, No Audio—Refusal)	Code 72 (Interview Complete, No Audio—Technical Problem)
1	13	2.50	2.50	13	0	0
2	111	21.39	23.89	108	2	1
3	101	19.46	43.35	96	5	0
4	73	14.07	57.42	66	5	2
5	43	8.29	65.70	39	3	1
6	43	8.29	73.99	38	4	1
7	29	5.59	79.58	25	3	1
8	17	3.28	82.85	16	1	0
9	17	3.28	86.13	16	0	1
10-15	49	9.44	95.57	47	0	2
16-20	18	3.47	99.04	17	1	0
21-25	3	0.58	99.61	2	1	0
26-30	2	0.39	100.00	1	0	1
31+	0	0.00	100.00	0	0	0
Total	519	100.00		484	25	10

MHSS = Mental Health Surveillance Study.

Table 7.14 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2010

Call Attempts (#)	Total Non-respondents (#)	Total Non-respondents (%)	Cumulative Total Non-respondents (%)	Code 73 (Breakoff, Partial Interview)	Code 74 (Unable to Contact)	Code 75 (Phone Number Problem)	Code 76 (Refusal)	Code 77 (Other)
1-5	11	8.94	8.94	7	2	0	1	1
6-10	8	6.50	15.45	0	5	0	1	2
11-15	33	26.83	42.28	5	23	0	1	4
16-20	25	20.33	62.60	1	22	1	1	0
21-25	32	26.02	88.62	2	28	1	0	1
26-30	13	10.57	99.19	0	11	1	1	0
31+	1	0.81	100.00	0	1	0	0	0
Total	123	100.00		15	92	3	5	8

MHSS = Mental Health Surveillance Study.

Table 7.15 shows the number of call attempts made to contact the MHSS respondents in 2011. The CIs completed more than 26 percent of the clinical interviews in one or two call attempts. Up to six call attempts were made to complete almost 75 percent of the clinical interviews. The CIs made 10 or more call attempts to complete 13 percent of the clinical interviews.

Table 7.16 shows the number of call attempts made to contact MHSS cases finalized as nonrespondents in 2011. For the vast majority of these cases, the CIs were unable to make contact with the case. The CIs determined 66 percent of these cases to be nonrespondents within 20 contacts, and almost 88 percent of interviews were classified as nonrespondents within 25 contacts.

Table 7.15 Number of Call Attempts Made To Contact MHSS Respondents: 2011

Call Attempts (#)	Total Respondents (#)	Total Respondents (%)	Cumulative Total Respondents (%)	Code 70 (Interview Complete, Audio Recorded)	Code 71 (Interview Complete, No Audio—Refusal)	Code 72 (Interview Complete, No Audio—Technical Problem)
1	66	4.36	4.36	61	3	2
2	335	22.13	26.49	319	13	3
3	288	19.02	45.51	269	12	7
4	220	14.53	60.04	212	6	2
5	136	8.98	69.02	130	5	1
6	85	5.61	74.64	77	4	4
7	84	5.55	80.13	80	1	3
8	49	3.24	83.42	45	4	0
9	51	3.37	86.79	47	1	3
10-15	134	8.85	95.64	133	1	0
16-20	48	3.17	98.81	47	1	0
21-25	11	0.73	99.54	9	2	0
26-30	7	0.46	100.00	5	2	0
31+	0	0.00	100.00	0	0	0
Total	1,514	100.00		1,434	55	25

MHSS = Mental Health Surveillance Study.

Table 7.16 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2011

Call Attempts (#)	Total Non-respondents (#)	Total Non-respondents (%)	Cumulative Total Non-respondents (%)	Code 73 (Breakoff, Partial Interview)	Code 74 (Unable to Contact)	Code 75 (Phone Number Problem)	Code 76 (Refusal)	Code 77 (Other)
1-5	27	7.09	7.09	10	0	7	4	6
6-10	41	10.76	17.85	12	10	12	4	3
11-15	82	21.52	39.37	14	56	10	0	2
16-20	102	26.77	66.14	14	82	5	0	1
21-25	82	21.52	87.66	8	71	1	0	2
26-30	37	9.71	97.38	2	35	0	0	0
31+	10	2.62	100.00	1	8	1	0	0
Total	381	100.00		61	262	36	8	14

MHSS = Mental Health Surveillance Study.

Table 7.17 shows the number of call attempts made to contact the MHSS respondents in 2012. Similar to previous years, about 25 percent of the clinical interviews were completed in one or two call attempts. The CIs made 10 or more call attempts to complete 13 percent of the clinical interviews, and more than 96 percent of cases were completed within 15 call attempts.

Table 7.18 shows the number of call attempts made to contact MHSS cases finalized nonrespondents. For the vast majority of these cases, the CIs were unable to make contact with the case. The CIs determined almost 60 percent of these cases to be nonrespondents within 20 contact attempts, and more than 83 percent of interviews were classified as nonrespondents within 25 contacts.

Table 7.17 Number of Call Attempts Made To Contact MHSS Respondents: 2012

Call Attempts (#)	Total Respondents (#)	Total Respondents (%)	Cumulative Total Respondents (%)	Code 70 (Interview Complete, Audio Recorded)	Code 71 (Interview Complete, No Audio—Refusal)	Code 72 (Interview Complete, No Audio—Technical Problem)
1	64	3.89	3.89	63	0	1
2	351	21.32	25.21	329	12	10
3	333	20.23	45.44	315	9	9
4	217	13.18	58.63	200	10	7
5	151	9.17	67.80	132	8	11
6	108	6.56	74.36	102	2	4
7	89	5.41	79.77	82	5	2
8	62	3.77	83.54	54	4	4
9	55	3.34	86.88	51	3	1
10-15	154	9.36	96.23	137	8	9
16-20	39	2.37	98.60	36	2	1
21-25	8	0.49	99.09	8	0	0
26-30	12	0.73	99.82	10	1	1
31+	3	0.18	100.00	3	0	0
Totals	1,646	100.00		1,522	64	60

MHSS = Mental Health Surveillance Study.

Table 7.18 Number of Call Attempts Made To Contact MHSS Nonrespondents: 2012

Call Attempts (#)	Total Non-respondents (#)	Total Non-respondents (%)	Cumulative Total Respondents (%)	Code 73 (Breakoff, Partial Interview)	Code 74 (Unable to Contact)	Code 75 (Phone Number Problem)	Code 76 (Refusal)	Code 77 (Other)
1-5	38	9.00	9.00	25	0	5	6	2
6-10	33	7.82	16.82	11	6	12	3	1
11-15	64	15.17	31.99	14	42	8	0	0
16-20	117	27.73	59.72	14	98	3	2	0
21-25	101	23.93	83.65	4	91	4	1	1
26-30	50	11.85	95.50	5	44	1	0	0
31+	19	4.50	100.00	2	17	0	0	0
Total	422	100.00		75	298	33	12	4

MHSS = Mental Health Surveillance Study.

7.3 Distressed Respondents

CIs were trained to fully assess signs and symptoms of emotional distress by the respondent during the administration of the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002) over the telephone. Respondents' distress included reports of recent suicidal or homicidal thoughts, plans, or actions, or showing strong feelings of sadness, irritability, or agitation during SCID administration. As shown in Appendix C, the MHSS had a detailed Distressed Respondent Protocol (DRP) to handle five different scenarios differentiated by risk of harm to self or others.

As summarized in Table 7.19, DRP Scenario #1 was for respondents reporting recent passive suicidal ideation, which included vague thoughts of suicide in the absence of a plan (risk of self-harm, no imminent danger). DRP Scenario #2 was for respondents reporting recent active suicidal ideation, which included specific plans for suicide and acting on those thoughts (risk of self-harm, possible/definite imminent danger). DRP Scenario #3 was for respondents reporting recent passive homicidal ideation, which included vague thoughts of homicide in the absence of a plan (risk of harm to others, no imminent danger). DRP Scenario #4 was for respondents who

reported recent active homicidal ideation, which included specific plans for homicide and acting on those thoughts (risk of harm to others, possible/definite imminent danger). DRP Scenario #5 was for respondents who showed signs of emotional distress during the SCID, such as sadness, irritability, or agitation, in the absence of suicidal or homicidal thoughts (no risk of harm, respondent is agitated or upset).

Table 7.19 Distressed Respondent Protocol Scenarios: 2008-2012 MHSS

Distressed Respondent Protocol Scenario	Risk of Harm	Imminent Danger
Scenario #1: Passive suicidal ideation	Self	No
Scenario #2: Active suicidal ideation	Self	Yes/Maybe
Scenario #3: Passive homicidal ideation	Other(s)	No
Scenario #4: Active homicidal ideation	Other(s)	Yes/Maybe
Scenario #5: Respondent agitated or upset	None	No

MHSS = Mental Health Surveillance Study.

As shown in Table 7.20, a total of 201 respondents were classified as distressed during the 2008-2012 MHSS. Among these, 145 were DRP Scenario #1 (passive suicidal ideation), 10 were Scenario #2 (active suicidal ideation), 4 were Scenario #3 (passive homicidal ideation), and 42 were Scenario #5 (respondent upset or agitated). There were no cases that were Scenario #4 (active homicidal ideation).

Table 7.20 Number of Distressed Respondent Protocol Classifications: 2008-2012 MHSS

Distressed Respondent Protocol Scenario	2008 (*n*=1,581[1])	2009 (*n*=547[1])	2010 (*n*=533[1])	2011 (*n*=1,569[1])	2012 (*n*=1,721[1])	2008-2012 (*n*=5,951[1])
Scenario #1: Passive suicidal ideation	22	9	13	50	51	145
Scenario #2: Active suicidal ideation	5	2	0	2	1	10
Scenario #3: Passive homicidal ideation	1	0	1	0	2	4
Scenario #4: Active homicidal ideation	0	0	0	0	0	0
Scenario #5: Respondent agitated or upset	4	0	2	10	26	42
Total	32	11	16	62	80	201

MHSS = Mental Health Surveillance Study.

[1] Total number of complete and incomplete Structured Clinical Interviews for DSM-IV.

CIs' adherence to the Institutional Review Board (IRB)–approved protocol for handling distressed respondents was 91.0 percent (n = 183) for the 2008-2012 MHSS. As summarized in Table 7.21, among the 18 cases involving protocol violations, 7 cases were Scenario #1 (passive suicidal ideation), 6 cases were Scenario #2 (active suicidal ideation), and 5 cases were Scenario #5 (respondent upset or agitated). Half of the protocol violations were due to unforeseen circumstances, such as trouble connecting the respondent with Lifeline services (n = 3), trouble connecting respondent with emergency services (n = 3), or lost phone connection before the CI was able to read the DRP script (n = 3). The remaining 9 cases resulted from CIs who did not recognize the respondent's distress at the time of the interview (n = 6), did not read the DRP script verbatim (n = 2), or recontacted the respondent after terminating the interview (n = 1). The distribution of IRB violations across DRP scenarios and study years is summarized in Table 7.22.

Table 7.21 Number of Institutional Review Board Violations, by Distressed Respondent Protocol Scenarios: 2008-2012 MHSS

Distressed Respondent Protocol Scenario	2008 (*n*=1,581[1])	2009 (*n*=547[1])	2010 (*n*=533[1])	2011 (*n*=1,569[1])	2012 (*n*=1,721[1])	2008-2012 (*n*=5,951[1])
Scenario #1: Passive suicidal ideation	1	1	1	2	2	7
Scenario #2: Active suicidal ideation	3	1	0	2	0	6
Scenario #3: Passive homicidal ideation	0	0	0	0	0	0
Scenario #4: Active homicidal ideation	0	0	0	0	0	0
Scenario #5: Respondent agitated or upset	0	0	1	3	1	5
Total	4	2	2	7	3	18

MHSS = Mental Health Surveillance Study.

[1] Total number of complete and incomplete Structured Clinical Interviews for DSM-IV.

Table 7.22 Number and Percentage of Cases with Institutional Review Board Violations, by Distressed Respondent Protocol Scenarios: 2008-2012 MHSS

Distressed Respondent Protocol Scenario	Total Distressed Respondent Protocol Cases (#)	Institutional Review Board Violations (#)	Cases with Institutional Review Board Violations (%)
Scenario #1: Passive suicidal ideation	145	7	4.83
Scenario #2: Active suicidal ideation	10	6	60.00
Scenario #3: Passive homicidal ideation	4	0	0.00
Scenario #4: Active homicidal ideation	0	0	0.00
Scenario #5: Respondent agitated or upset	42	5	11.90
Total	201	18	8.96

MHSS = Mental Health Surveillance Study.

To ensure that mental health services would be available should the respondent require them, modifications were made to the DRP in 2011. First, rather than having the CIs dial 9-1-1 (which routed the call to emergency services in the CI's area only), the CIs began using a national emergency number database to directly call emergency care providers nearest to the respondent. Second, rather than connecting respondents with Lifeline using three-way calling (which routed the call to Lifeline services in the CI's area only), the CIs asked the respondents to call Lifeline's national hotline directly. These modifications to the DRP reduced the likelihood of technical problems due to using three-way calling and increased the likelihood of respondents receiving local emergency and/or mental health services. To ensure that appropriate services were matched to level of risk, modifications to the DRP were made beginning in Quarter 2 of 2012 so that respondents reporting active suicidal ideation (Scenario #2) who denied intentions to harm themselves were referred to mental health counselors rather than being connected with emergency services.

Project management was notified of all 201 encounters with distressed respondents using electronic incident reports. The clinical supervisors (CSs) reviewed all of these incident reports, as well as the SCID booklets and audio files as needed. The IRB and members of the project management team were notified whenever emergency services were contacted as part of Scenario #2 (active suicidal ideation; $n = 10$) or Scenario #4 (active homicidal ideation; $n = 0$) and/or whenever there were DRP violations ($n = 18$, including the six cases involving

Scenario #2). In all 18 cases, the IRB was satisfied with the actions taken to protect the safety of the respondents. CIs who violated the DRP received retraining by the CSs.

7.4 Short Blessed Scale

As discussed in Section 4.2.4, a protocol was developed for encounters with respondents suspected of having problems with basic cognitive functions (such as attention, language production, orientation, language comprehension, and memory), which could lead to invalid data. As part of this Cognitive Impairment Protocol, the CIs were instructed to immediately break off the contact if the respondent showed signs of cognitive impairment during the introduction and informed consent process; however, if they had started the SCID, the CIs were to stop the interview and administer the Short Blessed Scale (SBS) to respondents. The SBS (Appendix C, page CIP.1) is a six-item scale designed to assess cognitive ability according to orientation, memory, and concentration. These questions are provided to CIs in the Cognitive Impairment Protocol section at the back of the SCID booklet. SBS scores indicate the number of errors, ranging from 0 to 28. Errors on the SBS can be indicative of temporary mental impairment (e.g., alcohol intoxication, medication side effects) or more long-term cognitive dysfunction (e.g., dementia, head injury).

For 10 or fewer errors on the SBS, CIs were instructed to resume the interview and to note the situation in the interviewer debriefing questions at the end of the SCID. For more than 10 errors, CIs were instructed to break off the interview and to document the situation when they entered the status code. CIs were instructed to attempt to complete the interview at another time if the respondent's cognitive impairment was deemed to be short term. If completing the SCID at another time, CIs were instructed to review the portions of the SCID completed earlier, to verify accuracy, and to include the reason for the breakoff in the debriefing questions.

As reported in Table 7.23, a total of 47 respondents exceeded the SBS cutoff score of 10 in the 2008-2012 MHSS. Their average SBS scores ranged from 15.2 to 18.0. More than half of those who exceeded the SBS cutoff score were females (57 percent; $n = 27$). The average age of respondents who exceeded the SBS cutoff score was 45. Long-term cognitive dysfunction was suspected in two thirds of the cases ($n = 31$), including developmental disabilities ($n = 20$) and self-reported cognitive disorders ($n = 11$). Short-term cognitive dysfunction may have been associated with the remaining third of the cases ($n = 16$) in which substance use ($n = 11$) and fatigue ($n = 5$) were suspected; however, none of these interviews were completed at another time. Data from these 47 respondents were considered invalid and not included in the final MHSS datasets or response rates.

Table 7.23 Demographic Characteristics of Respondents Who Exceeded the SBS Cutoff Score: 2008-2012 MHSS

Year	Cases	Average SBS Score	Gender	Average Age
2008	13	15.6	5 females, 8 males	46.1
2009	3	16.3	3 females, 0 males	47.5
2010	4	18.0	3 females, 1 males	43.8
2011	9	15.7	7 females, 2 males	35.4
2012	18	15.2	9 females, 9 males	48.9
Total	47	15.7*	27 females, 20 males	45.0*

MHSS = Mental Health Surveillance Study; SBS = Short Blessed Scale.

*Weighted average.

7.5 Problems Encountered

In Quarter 4 of 2011 (October 1-December 31, 2011), RTI data quality staff detected falsified work conducted on a large scale by one Pennsylvania (PA) field interviewer (FI). As a result, a total of 4,864 screenings and 3,412 interviews completed in PA and Maryland (MD) were removed from the 2006-2011 NSDUH data files. Because MHSS clinical interview respondents were recruited at the end of the main study interview beginning in 2008, it was possible for the PA FI to have falsified MHSS clinical interviews as well. Table 7.24 provides a summary by year of the number of respondents who were sampled for MHSS interviews from the main study respondents conducted by this PA FI. The table also includes the number of MHSS sampled cases who then agreed to complete the MHSS interview and the number of cases who completed the MHSS interview. Of the three completed MHSS clinical interviews, one clinical interview was completed in 2008 and two clinical interviews were completed in 2011.

Table 7.24 MHSS Interviews Associated with Main Study Interviews Completed, by Pennsylvania Field Interviewer: 2008-2011

Year	Selected for MHSS (#)	Agreed To Participate (#)	MHSS Interviews Completed (#)
2008	4	1	1
2009	3	0	0
2010	1	0	0
2011	9	4	2
Total	17	5	3

MHSS = Mental Health Surveillance Study.

Based on review of the SCID data collected, it was determined that all MHSS cases associated with the PA FI's main study interviews, that is, those that were selected for the MHSS clinical follow-up study, may have also been falsified and therefore removed from the final data files. Falsified cases were also removed for a MD FI during this same time period. Three of the MD FI's 2008 cases were selected for the MHSS, and one completed the MHSS interview.

The final result of the falsified cases from the PA and MD FIs was the removal of two interviews from the 2008 MHSS and two interviews from the 2011 MHSS. Because the PA FI's main study cases were dropped or reworked, her MHSS sampled nonrespondents were not considered as having been sampled for the MHSS clinical follow-up and consequently were not included in the nonresponse adjustments or data files. The MHSS weights from the affected years were recomputed to account for the removal of the PA and MD FIs' falsified interviews.

After processing the 2011 data, a field was added to the RTI internal consistency check Web site to identify MHSS cases linked to invalid main study cases. This new field enabled data quality staff to identify MHSS cases associated with invalid or falsified main study cases throughout the year. Once identified, the MHSS DCM assigned a code of "79: case reworked by FI, respondent not selected" to the MHSS case. This code indicated that the corresponding main study case was invalid for data quality purposes or falsification; thus, the MHSS case had been finalized as a nonrespondent.

8. Data Collection Quality Control

In addition to quality control measures implemented during recruiting, training, and certification of the clinical interviewers (CIs) (as discussed in Chapters 3 and 5) and the editing and keying of completed clinical interviews (as discussed in Chapter 6), this chapter further describes the procedures implemented during the Mental Health Surveillance Study (MHSS) data collection to maximize the quality and validity of the clinical interview data collected and delivered in the final analytic data file.

8.1 Clinical Supervision

The clinical supervisors (CSs) were responsible for monitoring the quality of the data collected by the CIs and providing timely feedback and retraining to the CIs during data collection. This quality control measure, referred to as "clinical supervision," helped to ensure that CIs were administering the Structured Clinical Interview for DSM-IV Axis I Disorders Non-patient Edition (SCID-I/NP or SCID; First, Spitzer, Gibbon, & Williams, 2002) according to MHSS protocol and using clinical interviewing best practices, while also providing CSs with the opportunity to review and identify invalid cases that should not be included in the final analytic data file.

As discussed in Section 6.3.1, once a completed SCID was received at RTI International, the CSs reviewed all of the data collected in every SCID booklet, item by item, comparing the notes provided by the CI with the diagnostic rating. The quality control protocol required that CIs receive full audio reviews of their first two recorded clinical interviews, along with at least a 10 percent random sample of subsequent recorded interviews. All audio-recorded interviews that were not selected for full review received partial reviews. For full reviews, both the audio recordings and the SCID booklet were reviewed in their entirety. If the interview was not selected for a full review, the CSs completed a partial review of the audio file(s), which included a review of the SCID's overview section and any other portions of the audio file(s) that needed review to clarify responses, to supplement notes, or to support the CI's ratings in the SCID booklet.

The results of the CS review of each clinical interview were documented on the SCID editing form (Appendix G). This form provided a record of clinical editing, a format for individualized feedback, and a system for monitoring and tracking CI performance. The CSs provided a rating of the CI's performance with the administration of each SCID module and documented the strengths and areas of improvement for the CI with clinical interviewing, including the following:

- providing accurate and sufficient notes,
- assigning proper or missed codes,
- amount of probing,
- types of probing,
- obtaining an adequate amount of information,

- obtaining the relevant information,
- resolving inconsistencies,
- maintaining a professional demeanor,
- establishing rapport with the respondent,
- dealing with the respondent's emotions properly,
- asking questions verbatim,
- obtaining a description of the experience in the respondent's own words,
- maintaining the correct time period,
- differentiating between symptoms that are easily confused,
- handling distressed respondents appropriately,
- assessing medical or substance etiologies, and
- completing the entire SCID following correct skip patterns.

An integral part of the data collection quality control process was providing CIs with constructive feedback from the CS reviews of the SCID booklets and audio files. This part of the process enabled the CIs to receive immediate feedback and retraining as needed to improve the quality of clinical interviews. Throughout the MHSS, CIs had individual (i.e., one-on-one) supervision teleconferences with CSs after each full review. These meetings involved in-depth discussion of the case reviewed by the CS and detailed feedback for the CI based on the ratings, strengths, and weaknesses noted on the SCID editing forms. The CSs also provided feedback on any partial reviews of SCID interviews conducted by the CI since the last individual CI supervision meeting, especially if there were any patterns of errors or poor quality across multiple interviews. During individual supervision meetings, CIs had the opportunity to ask questions and to ask for clarification on the feedback provided by the CSs. Also, supervision meetings included role-play exercises for demonstrating and practicing areas for improvement. For example, if the CI was not using adequate probing during the recorded interview, the CS would provide tips and suggestions on proper probes and then incorporate the need for probes into the mock exercise. As the CIs gained experience, these individual supervision meetings focused on honing and fine-tuning the CIs' clinical interviewing skills. The CSs tracked the number of clinical interviews completed by each CI and the dates of individual supervision to ensure that CIs received feedback on their performance in a timely manner (after completing their last interview). CIs met with a CS at least once per quarter to receive individual supervision and feedback on their performance.

To ensure timely supervision, CSs made themselves available by telephone and e-mail during regular business hours as well as evenings and weekends. This enabled CIs to contact CSs directly after conducting an interview to ask questions about a case or to debrief CSs on challenging cases. CIs were also required to seek supervision and notify a CS when the Distressed Respondent Protocol (DRP) was employed for any scenario. The CS also communicated immediately with a CI if a SCID arrived at RTI with a note on its cover indicating that it was a complex case. CSs immediately reviewed the SCID and audio files (as needed) for these cases.

As discussed in Section 5.4, to further ensure the quality of the data being collected in the clinical interviews, the CIs participated in mandatory end-of-quarter interrater reliability exercises. During these exercises, the CIs listened to a selected interview, independently rated the symptoms, and submitted their ratings to the CSs. The CIs' ratings were scored according to their agreement with the CS consensus ratings. After receiving their scored ratings, the CIs participated in a group conference call to discuss and calibrate their ratings. These exercises allowed the CSs to evaluate the CIs' diagnostic skills and to provide retraining as necessary.

8.2 Invalid Cases

Despite the quality control measures taken during MHSS data collection, clinical interviews were deemed invalid each year of the MHSS. Concerns with the quality and validity of these clinical interviews arose when either the respondent did not provide complete and accurate information or the CI did not adequately conceptualize the case or assess the symptoms. Completed interviews that one CS considered to be of questionable validity were subject to an independent second review by at least one other CS. If all CSs agreed that the interview did not meet quality standards, the case was deemed invalid (or unusable) and removed from the final analytic data file. This section describes the types of invalid clinical interviews on the MHSS, as well as how invalid main study cases affected the MHSS.

8.2.1 Invalid Clinical Interviews

As shown in Table 8.1, from 2008 to 2012, data for 298 cases (5.0 percent) were deemed unusable and were discarded from the dataset. Discarded cases included both completed and partial (or breakoff) interviews. The most common reason for a discarded interview was a respondent-induced breakoff, which accounted for 48.3 percent (n = 144) of all discarded interviews. Interviews in this category include cases in which respondents requested to terminate the interview (e.g., due to length, the personal nature of the questions) or ended (or appeared to end) the call without giving a reason. If the respondent requested to end the interview, the CI was instructed to inquire about the respondent's concerns and address problems if possible (e.g., reiterate confidentiality safety measures, schedule another time to finish the interview). If the respondent's concerns could not be satisfactorily addressed, the interview was terminated and no further contact was made. In the cases where the phone connection was lost, the CI made multiple attempts to reconnect with the respondent on the day of the initial breakoff and on future days. Interviews that were later completed were not discarded.

The second most common reason for a discarded interview was cognitive impairment. Among all discarded cases, 15.8 percent (n = 47) were due to cognitive impairment as captured by the Short Blessed Scale (SBS). CIs were instructed to administer the SBS to respondents who were suspected of cognitive impairment. Respondents who had more than 10 errors on the SBS were classified as having cognitive impairment. Once a respondent was determined to have cognitive impairment based on his or her score on the SBS, the CI terminated the interview and proceeded directly into the End of Interview script. (See Section 7.4 for more information on the SBS.)

Another 6.4 percent (n = 19) of all discarded cases were with other respondents who demonstrated signs of cognitive impairment significant enough to affect data quality but who either obtained a passing score on the SBS (less than or equal to 10 errors) or who were not

administered the SBS. The CSs reviewed these cases and determined whether the suspected cognitive impairment significantly reduced the validity of the data and should therefore be discarded.

Language comprehension problems were responsible for 3.0 percent (n = 9) of all discarded cases and mainly occurred because English was the respondent's second language. The MHSS interview was conducted in English only. These cases were generally completed interviews (not breakoffs) in which the respondent reported speaking English as a second language, exhibited comprehension problems that persisted even after the CI rephrased or clarified questions, and either answered "no" to nearly all items or provided answers that demonstrated a lack of understanding of the question. Cases in which the CS determined that the respondent appeared to understand the question after the CI rephrased or clarified questions were not discarded.

Interviews terminated due to respondent distress accounted for 7.4 percent (n = 22) of all discarded cases. When a respondent became distressed during the interview, the CI used his or her clinical judgment to determine the appropriate course of action: (1) continue the interview after checking in with the respondent, (2) activate the appropriate DRP scenario and terminate the interview with no further contact, or (3) activate the appropriate DRP scenario and schedule a time to continue the interview later (if the respondent was willing and if it did not appear that finishing the interview would be detrimental to the respondent). Interviews that the respondent was willing and able to continue or complete at a later date were not discarded.

Only 4.0 percent of the discarded cases (n = 12) were discarded because of CI error. Most cases that were lost because of CI error had complicated clinical presentations and occurred when the CI's incorrect clinical ratings led to missing data, the CI failed to collect sufficient data, or the CI failed to resolve inconsistencies in the respondent's reports.

A final category for discarded cases was used for cases of questionable validity that the other categories did not capture. Among all the discarded cases, 14.4 percent (n = 43) fell into this category of questionable data. This category included respondents who (1) were unwilling to disclose personal information or provide detailed responses, (2) provided inconsistent information that could not be resolved, (3) appeared to answer questions haphazardly or not take the interview seriously, (4) grossly over- or under-endorsed symptoms, or (5) had complicated clinical presentations that the MHSS could not fully assess. Also, two cases were coded as "Other" due to quality issues with the corresponding main study interview.

The CS team thoroughly reviewed each case that was considered for discarding and closely reviewed the audio recording. As described in Section 8.1, CIs were provided with individual feedback and retraining if their actions might have contributed to a breakoff or compromised data quality. As described in Section 5.3, CSs also conducted group retraining sessions with all veteran CIs to review common data quality errors and best practices for clinical interviewing. These sessions included example scenarios from actual clinical interviews for discussion during the conference call. The goals of these retraining sessions were to maximize the quality of the data collected in the field and to decrease the likelihood of discarded data in the future.

Table 8.1 Types of Discarded Cases, by Year: 2008-2012 MHSS

Code	Description	2008 (*n*=1,581[1]) Cases (#)	2009 (*n*=547[1]) Cases (#)	2010 (*n*=533[1]) Cases (#)	2011 (*n*=1,569[1]) Cases (#)	2012 (*n*=1,721[1]) Cases (#)	2008-2012 (*n*=5,951[1]) Cases (#)	Total (%)
90.1	Respondent-induced	38	14	9	42	41	144	48.3
90.2	Language comprehension	4	1	1	2	1	9	3.0
90.3	Respondent distress	7	1	0	3	11	22	7.4
90.4	Cognitive (fail SBS)	13	3	4	9	18	47	15.8
90.5	Cognitive (non-SBS)	3	1	0	6	9	19	6.4
90.6	Questionable data	14	7	2	6	14	43	14.4
90.7	Clinical interviewer error	1	0	1	6	4	12	4.0
	Other	1	0	0	0	1	2	0.7
	Total	81	27	17	74	99	298	
	Percentage of total number of attempted SCID interviews[1]	5.1%	4.9%	3.2%	4.7%	5.8%	5.0%	

MHSS = Mental Health Surveillance Study; SBS = Short Blessed Scale; SCID = Structured Clinical Interview for DSM-IV.

[1] Total number of complete and incomplete SCID interviews.

8.2.2 Invalid Main Study Interviews

As part of the National Survey on Drug Use and Health (NSDUH) main study data quality procedures, project staff perform ongoing mail, telephone, and field verifications of field interviewer work throughout the year. During this process, some interviews are found to have been completed at the wrong dwelling unit, conducted with the incorrect respondent, conducted with a serious protocol violation (e.g., conducted without parental consent), or falsified by an interviewer. These NSDUH main study interviews are considered invalid and are removed from the dataset. If the invalid main study interview had yielded an MHSS interview, the CI was notified immediately not to conduct the MHSS interview. If the MHSS interview was already completed, it was considered invalid and removed from the dataset.

Main study interviews that did not pass the "usable case rule" also were considered invalid. The usable case rule requires that a NSDUH respondent answer "yes" or "no" to the question on lifetime use of cigarettes and "yes" or "no" to at least nine additional lifetime use questions in the core NSDUH modules. If the NSDUH interview did not pass the usable case rule, it and the corresponding MHSS interview were both removed from the final dataset. These invalid NSDUH and MHSS interviews were identified and removed from the final dataset after completion of data collection for the entire survey year. See the 2011 NSDUH mental health findings report (Center for Behavioral Health Statistics and Quality, 2012) for more information on the usable case rule as it is applied to the main NSDUH interview.

References

Aldworth, J., Barnett-Walker, K., Chromy, J., Karg, R., Morton, K., Novak, S., & Spagnola, K. (2009, June). *Measuring serious mental illness with the NSDUH: Results of 2008 12-month analysis* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. 283-2004-00022, Mental Health Surveillance Study Deliverable No. 5, RTI/0209009.423.006.008). Research Triangle Park, NC: RTI International.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington, DC: American Psychiatric Association.

American Psychiatric Publishing. (2000). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision* (DSM-IV-TR; doi: 10.1176/appi.books.9780890423349). Retrieved from http://www.psychiatryonline.com/resourceTOC.aspx?resourceID=1

Basco, M. R., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*, 1599-1605.

Center for Behavioral Health Statistics and Quality. (2012). *Results from the 2011 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 12-4725, NSDUH Series H-45). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Crippa, J. A., de Lima Osório, F., Del-Ben, C. M., Filho, A. S., da Silva Freitas, M. C., Loureiro, S. R. (2008, October). Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. *Perspectives in Psychiatric Care, 44*(4), 241-247.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*, 766-771.

First, M. B., Spitzer R. L., Gibbon M., & Williams J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research Department.

Hajebi, A., Motevalian, A., Amin-Esmaeili, M., Hefazi, M., Radgoodarzi, R., Rahimi-Movaghar, A., & Sharifi, V. (2012, July). Telephone versus face-to-face administration of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, for diagnosis of psychotic disorders. *Comprehensive Psychiatry, 53*(5), 579-583.

Jackson, J. S., Neighbors, H. W., Nesse, R. M., Trierweller, S. J., & Torres, M. (2004). Methodological innovations in the National Survey of American Life. *International Journal of Methods in Psychiatric Research, 13*, 289-298.

Jordan, B. K., Karg, R. S., Batts, K. R., Epstein, J. F., & Wiesen, C. A. (2008). A clinical validation of the National Survey on Drug Use and Health Assessment of Substance Use Disorders. *Addictive Behaviors, 33*, 782-798.

Kessler, R., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research, 13*(2), 122-139.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S-L. T., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184–189.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

Lee, S., Tsang, A., Lau, L., Mak, A., Ng, K. L., & Chan, D. M. (2008, December). Concordance between telephone survey classification and face-to-face structured clinical interview in the diagnosis of generalized anxiety disorder in Hong Kong. *Journal of Anxiety Disorders, 22*(8), 1403-1411.

Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine, 27*(2), 93-105.

Liao, D., Kott, P., Aldworth, J., Yu, F., Karg, R., Shook-Sa, B., & Davis, T. (in press). *2012 Mental Health Surveillance Study: Design and estimation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283201000003C, RTI 0212800.001.112.008.006). Research Triangle Park, NC: RTI International.

Liao, D., Kott, P., Aldworth, J., Yu, F., Karg, R., Shook-Sa, B., & Davis, T. (2012). *2011 Mental Health Surveillance Study: Design and estimation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, RTI 0211838.212.008). Research Triangle Park, NC: RTI International.

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research, 19*(Suppl. 1), 49-60. doi:10.1002/mpr.313

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*(2), 110-123. doi:10.1002/mpr.61

Zanarini, M. C., Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., Morey, L. C., Grilo, C. M., Shea, M. T., McGlashan, T. H., & Gunderson, J. G. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders, 14*, 291-299.

Appendix A: CAI Recruitment Scripts

Forms Answer Navigate Options

You have been randomly selected to participate in one additional study for the U. S. Public Health Service. This interview will ask questions about mental health issues. It will be conducted over the telephone and will take about an hour. Participation in this interview is voluntary and all of your answers will be kept confidential.

HAND FOLLOW-UP STUDY DESCRIPTION TO RESPONDENT. Please read this statement. It describes the survey and the legislation that assures the confidentiality of any information you provide.

If you agree to participate, I will pay you an additional $30 today. Within the next two weeks, a different interviewer will call you to explain more about the interview and to schedule a convenient time to complete it. If you wish, you may complete the full interview when the interviewer calls.

IF ASKED "WHY WAS I SELECTED?" (You have been randomly chosen for this special study for the National Survey on Drug Use and Health. This study, sponsored by the U.S. Public Health Service, will ask questions about various mental health issues. Knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.)

- 1. RESPONDENT AGREES TO RECONTACT
- 2. RESPONDENT DOES NOT AGREE TO RECONTACT
- 3. RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD

RECRUIT1

Forms Answer Navigate Options

Since another interviewer will be completing the second interview, may I have your first name and phone number so the interviewer can call you?

ENTER FIRST NAME ONLY AND PHONE NUMBER.

To check that I entered the number correctly, please repeat the phone number.

CONFIRM NUMBER. AS NEEDED, READ THE CONTACT INFORMATION ENTERED TO THE RESPONDENT AND CONFIRM IT IS CORRECT.

Is there another number where the telephone interviewer could contact you about the second interview?

IF YES: RECORD PHONE NUMBER AND TYPE (CELL, WORK, ETC) IN THE NOTES FIELD.
REPEAT ABOVE STEPS TO CONFIRM THE NUMBER. YOU MAY ENTER UP TO 50 CHARACTERS.

IF NO: CONTINUE.

PRESS [ENTER] TO CONTINUE.

RECRUIT2 - FIRST NAME

PHONE NUMBER (___)___-____

NOTES

Forms Answer Navigate Options

Please also let me know the best days and times when you will be available in the next two weeks. I will give this information to the interviewer, and he or she will try to contact you during one of these times.

ENTER BEST DAYS/TIMES. AS NEEDED, PROBE FOR ADDITIONAL BEST DAYS/TIMES.
READ THE INFORMATION ENTERED TO THE RESPONDENT AND CONFIRM IT IS CORRECT.

COMPLETE A REMINDER CARD AND HAND TO THE RESPONDENT.

I have entered these days and times in the computer and recorded them on this card. Please note the interviewer may try to reach you at other times as well.

INTERVIEWER NOTE: ADDITIONAL INFORMATION REGARDING THE BEST DAYS OR TIMES PROVIDED BY THE RESPONDENT SHOULD BE ENTERED IN THE NOTES FIELD. YOU MAY ENTER UP TO 50 CHARACTERS.

TELEPHONE INTERVIEWERS ARE AVAILABLE EVENINGS AND WEEKENDS.

PRESS [ENTER] TO CONTINUE.

RECRUIT3 - BEST DAYS	
BEST TIMES	
NOTES	

Forms Answer Navigate Options

PAY RESPONDENT $30 CASH.

SIGN FOLLOW-UP INTERVIEW PAYMENT RECEIPT FORM AND GIVE TOP COPY TO RESPONDENT.

I have signed this form to indicate that I have paid you the $30 for the telephone interview.

[IF THE RESPONDENT WILL NOT ACCEPT THE CASH INCENTIVE, MARK THE APPROPRIATE BOX ON THE FOLLOW-UP INTERVIEW PAYMENT RECEIPT FORM.]

PRESS [ENTER] TO CONTINUE.

INCENTMH	

Forms Answer Navigate Options

Thank you for your time. I just need to finish a few questions on my own to show that I did the interview. This will only take a few minutes.

BE SURE YOU HAVE YOUR SHOWCARD BOOKLET, CALENDAR, QC ENVELOPE W/ FORM, AND PAYMENT RECEIPT COPIES.

PRESS [ENTER] TO CONTINUE.

THANKR2

If unavailable during the next two weeks (response 2 on RECRUIT1):

Forms Answer Navigate Options

To accommodate your schedule, an interviewer will be available to call you about this study and schedule a convenient time to complete the interview within the next four weeks.

- 1. RESPONDENT AGREES TO RECONTACT
- 2. RESPONDENT DOES NOT AGREE TO RECONTACT
- 3. RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD

RECRT4WK

If available during the next four weeks:

Forms Answer Navigate Options

Please also let me know the best days and times when you will be available in the next four weeks. I will give this information to the interviewer, and he or she will try to contact you during one of these times.

ENTER BEST DAYS/TIMES. AS NEEDED, PROBE FOR ADDITIONAL BEST DAYS/TIMES.
READ THE INFORMATION ENTERED TO THE RESPONDENT AND CONFIRM IT IS CORRECT.

COMPLETE A REMINDER CARD AND HAND TO THE RESPONDENT.

I have entered these days and times in the computer and recorded them on this card. Please note the interviewer may try to reach you at other times as well.

INTERVIEWER NOTE: ADDITIONAL INFORMATION REGARDING THE BEST DAYS OR TIMES PROVIDED BY THE RESPONDENT SHOULD BE ENTERED IN THE NOTES FIELD. YOU MAY ENTER UP TO 50 CHARACTERS.

TELEPHONE INTERVIEWERS ARE AVAILABLE EVENINGS AND WEEKENDS.

PRESS [ENTER] TO CONTINUE.

RECRUIT3 - BEST DAYS	
BEST TIMES	
NOTES	

If the respondent does not agree to participate (response 2 on RECRUIT1 or RECRT4WK):

Forms Answer Navigate Options

Thank you for your time. I just need to finish a few questions on my own to show that I did the interview. This will only take a few minutes.

BE SURE YOU HAVE YOUR SHOWCARD BOOKLET, CALENDAR, QC ENVELOPE W/ FORM, AND PAYMENT RECEIPT COPIES.

PRESS [ENTER] TO CONTINUE.

THANKR2

Appendix B: Recruitment Materials

Exhibit B.1 Follow-up Study Description

U.S. Public Health Service
Center for Behavioral Health
Statistics and Quality

Follow-up Study Description

You have been randomly chosen for this special study for the 2012 National Survey on Drug Use and Health. This study, sponsored by the United States Public Health Service, will ask questions about various mental health issues such as depression, anxiety, post traumatic stress disorder and substance dependence. Although there is no benefit to you personally, knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.

If you agree to participate in this follow-up interview, your first name and telephone number will be collected but will be used only for re-contact purposes. Your name and telephone number will not be included on the interview forms on which your answers will be written, or on any interview audio files that might be recorded. While the interview has some personal questions, federal law protects the privacy of your answers and requires us to keep all of your answers confidential. Any data that you provide will only be used by authorized personnel for statistical purposes according to the Confidential Information Protection and Statistical Efficiency Act of 2002. The only exception to this promise of confidentiality is if you tell the interviewer that you intend to seriously harm yourself or someone else; in this situation RTI may need to notify a mental health professional or other authorities.

The interview will be conducted over the phone and takes on average an hour to complete. Your participation is voluntary. You may consider some of the questions to be sensitive in nature and some of the questions also may make you feel certain emotions, such as sadness. Remember that you can refuse to answer any questions that you do not want to answer, and you can stop the interview at any time. If you become upset at any time during the interview and wish to speak to a mental health professional about how you are feeling, the interviewer can again provide you with the toll-free hotline numbers that are printed on your payment receipt from the first interview. **If you agree to complete the interview, you will receive $30 today.**

If you have questions about the study, call the Project Representative at . If you have questions about your rights as a study participant, call RTI's Office of Research Protection at (a toll-free number). You can also visit our project Website: http://nsduhweb.rti.org/ for more information.

Thank you for your cooperation and time.

, Project Officer
Center for Behavioral Health Statistics and Quality
Substance Abuse and Mental Health Services Administration (SAMHSA)
U.S. Public Health Service
Department of Health and Human Services

Exhibit B.2 Follow-up Interview Payment Receipt

Follow-up Interview Payment Receipt

United States Public Health Service

and

Research Triangle Institute

thank you for agreeing to participate in a special study for the 2012 National Survey on Drug Use and Health.

In appreciation of your participation in this important study, you are eligible to receive a $30 cash payment.

Since maintaining the confidentiality of your information is important to us, your name will not be entered on this form. However, the interviewer must sign and date this form to certify you received (or declined) the cash payment.

_____	_____	__ __ __ __ __ __ __ __ __ __ - __
Interviewer	Date	Case ID

☐ Accepted Cash Payment ☐ Declined Cash Payment

If you ever feel that you need to talk to someone about mental health issues, you can call the National Lifeline Network. Counselors are available to talk at any time of the day or night and they can give you information about services in your area.

1-800-273-TALK or 1-800-273-8255
1-888-628-9454 (Spanish)
http://suicidepreventionlifeline.org/

If you ever feel that you need to talk to someone about drug use issues, you can call the Substance Abuse and Mental Health Services Administration's Treatment Referral Helpline. This is a 24-hour service that will help you locate treatment options near you.

1-800-662-HELP or 1-800-662-4357
1-800-487-4889 (TDD)
http://findtreatment.samhsa.gov

Disposition: Top copy to Respondent, yellow to Field Supervisor, pink to Field Interviewer.

Exhibit B.3 Reminder Card

We appreciate you taking time for this important study and look forward to speaking with you soon.

Your suggested contact days and times are:

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Time							

Research Triangle Institute
Research Triangle Park, NC 27709-2194

Appendix C: Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I)

OMB No.: 0930-0110
Expiration Date: 08-31-14
V.8, 03-06-12

STRUCTURED CLINICAL INTERVIEW FOR DSM-IV AXIS I DISORDERS (SCID-I)

By

Michael B. First, M.D.; Miriam Gibbon, M.S.W.;

Robert L. Spitzer, M.D.; and Janet B. W. Williams, D.S.W.

MODIFIED BY RTI INTERNATIONAL

FOR

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

MENTAL HEALTH SURVEILLANCE STUDY

SCID Transmittal Record			
Interviewer ID:		QuestID:	
		Date of Interview: _____/ _____/ _____ MM DD YY	
Date Shipped to RTI:	_____/ _____/ _____ MM DD YY	Date Received at RTI:	_____/ _____/ _____ MM DD YY
Clinical QC by:		Date of Clinical QC:	_____/ _____/ _____ MM DD YY
Edited by:		Date Edited:	_____/ _____/ _____ MM DD YY

Public reporting burden for this collection of information is estimated to 60 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any aspect of this collection of information, including suggestions for reducing this burden to SAMHSA Reports Clearance Officer; Paperwork Reduction Project (0930-0110); Room 8-1099, 1 Choke Cherry Road, Rockville, MD 20857. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this project is 0930-0110.

Introduction to Clinical Interview

Before you call, **be prepared**:

- Review the assignment information provided including the respondent name, telephone number, as well as the date of the initial interview.
- Have your schedule available (in case you need to schedule an appointment).
- Have all interviewing materials available.

VERIFY NUMBER AND LOCATE RESPONDENT

Hi, my name is _____ and I'm calling on behalf of the U.S. Public Health Service. Is this [PHONE NUMBER]?

YES: PROCEED BELOW
NO: I apologize. I need to double check my records. Thank you for your time. END CALL.

I'm trying to reach [FIRST NAME] who agreed to take part in a telephone interview we're conducting. May I speak to [FIRST NAME]?

IF R NOT HOME OR UNAVAILABLE

When would be a good time to call again? ENTER CODE 51 AND DETAILS IN CMS. Thank you for your time. END CALL.

IF R AVAILABLE

(Hi, my name is _____.)
You recently completed an interview in your home with an interviewer working on the National Survey on Drug Use and Health. I am the interviewer you were told would contact you for a follow-up telephone interview. Do you recall completing the first interview?

YES: PROCEED BELOW.
NO: VERIFY FIRST NAME OF PERSON YOU ARE SPEAKING TO.
IF NOT SPEAKING TO CORRECT RESPONDENT, ASK TO SPEAK TO RESPONDENT.

IF NAME IS CORRECT AND RESPONDENT DOESN'T RECALL INITIAL INTERVIEW, REMIND OF DATE OF INITIAL INTERVIEW.

IF CORRECT RESPONDENT STILL NOT FOUND: I apologize. I need to double check my records. Thank you for your time. END CALL. ENTER CODE 59 AND INVESTIGATE.

Are you in a place where you can safely talk on the phone and answer my questions?

YES: PROCEED
NO: Are you able to move to a place where you can safely talk?
YES: PAUSE, THEN CONTINUE
NO: When would be a good time to call again? ENTER CODE 50 AND DETAILS IN CMS. Thank you for your time. END CALL.

Is now a good time to complete this interview?

YES: PROCEED. BE SURE TO READ VERBATIM.
NO: When would be a good time to call again? ENTER CODE 50 AND DETAILS IN CMS. Thank you for your time. END CALL.

PRIVACY

Because you may not want others to hear the responses to some of our questions, I'd like to be sure you're in a private area. Where are you right now? Are you at home, at work, or somewhere else? Are you in an area where you can answer these questions privately?

YES: PROCEED
NO: Please consider moving to a more private area. Do you need more time?
YES: PAUSE, THEN CONTINUE
NO: CONTINUE

INFORMED CONSENT

Before we begin, I would like to remind you of the study details. This study, sponsored by the United States Public Health Service, asks questions about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Although there is no benefit to you personally, knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States. While the interview has some personal questions, federal law keeps your answers private. The only exception to this promise of confidentiality is if you tell me that you intend to seriously harm yourself or someone else; in this situation I may need to notify a mental health professional or other authorities.

Your participation is voluntary. You may consider some of the questions to be sensitive in nature and some of the questions may also make you feel certain emotions, such as sadness. Remember that you can refuse to answer any questions that you do not want to answer, and you can stop the interview at any time. If you become upset at any time during the interview and wish to speak to a mental health professional about how you are feeling, I will provide you with the toll-free hotline numbers that are printed on your payment receipt from the first interview. It is important for you to keep in mind that I will not be providing you with a psychological diagnosis or any mental health advice or counseling. The information we are collecting today is only for research purposes.

These study details are also included on the Follow-up Study Description you received from the interviewer who met with you in your home. Do you have any questions before we begin?
ANSWER ANY RESPONDENT QUESTIONS.

Is it OK to continue with the interview?

YES: PROCEED TO NEXT PAGE
NO: BASED ON CONVERSATION:
What sort of concerns do you have about participating?
OR
Are there other questions that I could answer for you?

IF R STILL UNWILLING TO PARTICIPATE: Thank you for your time. END CALL.

RECORDING PERMISSION

In order to ensure that I am conducting this interview accurately and properly, I would like to make an electronic audio recording of this interview. This is done strictly for quality control purposes. The recording will only be listened to by staff members on the project who have signed confidentiality pledges. The recording will be stored in a secure manner and will not contain your name—only a random number that will be assigned to this case. To help maintain confidentiality, we ask that you not give your name or any other identifying information, such as an address or place of business, during the interview. All recordings will be permanently destroyed within eighteen months after the end of the data collection period. You can still do the interview if you do not want me to record it.

Do you agree to allow me to record the interview?

YES: I will now begin recording. START RECORDING AND SAY: "This is [YOUR FIRST AND LAST NAME] conducting telephone interview [QUEST ID] on [DATE]."
NO: DON'T RECORD

Ok, let's get started.

CI NOTES:
IF ASKED AT ANY TIME BY A RESPONDENT WHETHER THE INTERVIEWER IS A DOCTOR, PSYCHIATRIST, PSYCHOLOGIST, SOCIAL WORKER, OR OTHER MENTAL HEALTH PROFESSIONAL, YOU MAY DISCLOSE THAT YOU HAVE MEDICAL OR PSYCHOLOGICAL TRAINING THAT ALLOWS YOU TO FULLY UNDERSTAND THE SURVEY.

HOWEVER, YOU SHOULD EXPLAIN THAT YOUR INVOLVEMENT IN THIS STUDY IS FOR RESEARCH PURPOSES ONLY AND IN NO WAY CONSTITUTES MEDICAL OR PSYCHOLOGICAL ADVICE, TREATMENT, OR DIAGNOSIS. EXPLAIN THAT THIS IS NOT THE NATURE OF THIS EFFORT.

IF RESPONDENT REQUESTS PSYCHOLOGICAL COUNSELING OR ADVICE OF ANY KIND, REFER HIM/HER TO THE NATIONAL LIFELINE. IF RESPONDENT IS INTERESTED IN CONTACTING THE LIFELINE, OFFER TO STAY ON THE PHONE AND CONNECT THEM VIA A THREE-WAY CALL.

This page has been intentionally left blank.

OVERVIEW

Interview Start Time: ___ ___ : ___ ___ *AM/PM*

I'm going to be asking you about problems or difficulties you may have had, and I'll be making some notes as we go along.

DEMOGRAPHIC DATA

SEX: 1 male
2 female

OV1

What's your date of birth?

DOB: _____/ _____/ _____
mm dd yyyy

OV2

Are you married?

IF NO: Have you ever been married?

MARITAL STATUS (most recent):
1 married or living with someone as if married
2 widowed
3 divorced or annulled
4 separated
5 never married

OV3

Do you have any children?

IF YES: How many?
(What are their ages?)

_____

Where do you live?
(That is, do you live in a house, an apartment, or do you have some other living arrangement?)

_____

_____

Who do you live with?
(Do you live with family, friends, or roommates?)

EDUCATION AND WORK HISTORY

OV4

What's the highest grade or year of school you have completed?

EDUCATION:
1 grade 6 or less
2 grade 7 to 12 (without graduating high school)
3 graduated high school or high school equivalent
4 part college
5 graduated 2 year college
6 graduated 4 year college
7 part graduate/professional school
8 completed graduate/ professional school

IF FAILED TO COMPLETE A PROGRAM IN WHICH THEY WERE ENROLLED: Why did you decide to leave school?

_____

_____

What kind of work do you do?
(Do you work outside of your home?)

_____

_____

Are you working now?

IF YES: How long have you worked there? _____

IF LESS THAN 6 MONTHS: Why did you leave your last job? _____

Have you always done that kind of work? _____

IF NO: Why is that? How long has it been since you worked outside the home? What kind of work have you done? _____

How are you supporting yourself now? _____

IF UNKNOWN: Has there ever been a period of time when you were unable to work or go to school? _____

IF YES: Why was that? _____

PAST PERIODS OF PSYCHOPATHOLOGY

(THE LIFE CHART ON PAGE VIII OF OVERVIEW MAY BE USED TO SUMMARIZE A COMPLICATED HISTORY OF PSYCHOPATHOLOGY AND TREATMENT.)

Have you ever seen anybody for emotional or psychiatric problems?

Treatment for emotional problems with a physician or mental health professional — 1 NO, 3 YES — OV5

IF YES: What was that for? (What treatment(s) did you get? Any medications?) _____

If OV5= 1, SKIP OV5a

IF NO: Was there ever a time when you received medication to help your mood, calm your nerves, or to help you sleep? _____

IF NO: Was there ever a time when you, or someone else, thought you should see someone because of the way you were feeling or acting? _____

IF NOT ALREADY KNOWN:

Did you receive any of the treatment you just mentioned in the past 12 months, that is since (this date), 2011?

Most recent mental health treatment — OV5a

1 Lifetime MH Treatment (not Past Year)
2 Past Year – Counseling Alone
3 Past Year - Meds alone
4 Past Year – Counseling and meds

NOTE: IF OV5a = 3 or 4, medi-cations should be listed at OV7a or OV7b

What about treatment for drugs or alcohol? _____

Have you ever been a patient in a psychiatric hospital?

IF YES: What was that for? (How many times?)

IF GIVES AN INADEQUATE ANSWER, CHALLENGE GENTLY: e.g., Wasn't there something else? People don't usually go to psychiatric hospitals just because they are tired or nervous.

Number of previous hospitalizations (Do not include transfers) 0 1 2 3 4 5 (or more) — OV6

IF OV6 = 0, SKIP OV6a and OV6b

Timing of most recent psychiatric hospitalization — OV6a

1 Lifetime Psychiatric Hospitalizations (not Past Year)

2 Past Year Psychiatric Hospitalizations

IF OV6a = 1, SKIP OV6b

Reasons for hospitalization in the past year — OV6b

Have you ever been in a hospital for treatment of a medical problem?

IF YES: What was that for?

Thinking back over your whole life, when were you the most upset?

(Why? What was that like? How were you feeling?)

When were you feeling the best you have ever felt?

PSYCHOPATHOLOGY DURING PAST YEAR

Now I would like to ask you about the past year, that is since (CURRENT DATE) 2011. How have things been going for you?

Has anything happened that has been especially hard for you?

What about difficulties at work or with your family?

How has your mood been?

How has your physical health been? (Have you had any medical problems?) (USE THIS INFORMATION TO CODE AXIS III)

Thinking back over the past year, when were you the most upset?

(IF UNKNOWN:) Are you currently in a relationship?

IF YES: Tell me a little about that.

IF NO: How long has it been since you were in a relationship?

Do you take any medications or vitamins?

IF YES: How much and how often do you take (MEDICATION)? (What is that medication for?) (Has there been any change in the amount you have been taking?)

Are there any medications that you have taken in the past year that you are not currently taking?

Psychotropic medications taken in the past year (but not currently) OV7a

Psychotropic medications taken currently OV7b

How much have you been drinking (alcohol) (in the past year)?

Have you been taking any drugs (in the past year)? (What about marijuana, cocaine, other street drugs?)

Have you (in the past year) gotten “hooked” on a prescribed medicine or taken a lot more of it than you were supposed to?

CURRENT SOCIAL FUNCTIONING

How have you been spending your free time?

Who do you spend time with?

MOST LIKELY CURRENT DIAGNOSES:

DIAGNOSES THAT NEED TO BE RULED OUT:

LIFE CHART

Age (or date)	Description (symptoms, triggering events)	Treatment

RETURN TO OVERVIEW PAGE iv, **PSYCHOPATHOLOGY DURING PAST YEAR.**

SCID SCREENING MODULE

Now I want to ask you some more specific questions about problems you may have had. We'll go into more detail about them later.

RESPOND TO POSITIVE RESPONSES WITH: We'll talk more about that later.

1. In the past year, that is since (CURRENT DATE) 2011, have you had any panic attacks, when you suddenly felt frightened or anxious or suddenly developed a lot of physical symptoms?

1	2	3	
CIRCLE "1" ON PAGE E.9		CIRCLE "3" ON PAGE E.9	S1

2. In the past year, have you been afraid of going out of the house alone, being alone, being in a crowd, standing in a line, or traveling on buses or trains?

1	2	3	
CIRCLE "1" ON PAGE E.15		CIRCLE "3" ON PAGE E.15	S2

3. During the past year, has there been anything that you have been afraid to do or felt uncomfortable doing in front of other people, like speaking, eating, or writing?

1	2	3	
CIRCLE "1" ON PAGE E.19		CIRCLE "3" ON PAGE E.19	S3

4. In the past year have there been any other things that you've been especially afraid of, like flying, seeing blood, getting a shot, heights, closed places, or certain kinds of animals or insects?

1	2	3	
CIRCLE "1" ON PAGE E.23		CIRCLE "3" ON PAGE E.23	S4

5. In the past year have you been bothered by thoughts that didn't make any sense and kept coming back to you even when you tried not to have them?

   IF NOT SURE WHAT IS MEANT: Thoughts like hurting someone even though you really didn't want to or being contaminated by germs or dirt.

1	2	3	
CIRCLE "1" ON PAGE E.27		CIRCLE "3" ON PAGE E.27	S5

S6

6. In the past year has there been anything that you had to do over and over again and couldn't resist doing, like washing your hands again and again, counting up to a certain number, or checking something several times to make sure that you'd done it right?

1	2	3
CIRCLE "1" ON PAGE E.29		CIRCLE "3" ON PAGE E.29

7. In the past year, have you had times when you have been particularly nervous or anxious?

1	2	3	
CIRCLE "1" ON PAGE E.33		CIRCLE "3" ON PAGE E.33	S7

8. During the past year, have you had a time when you weighed much less than other people thought you ought to weigh?

1	2	3	
CIRCLE "1" ON PAGE F.1		CIRCLE "3" ON PAGE F.1	S8

9. In the past year, have you often had times when your eating was out of control?

1	2	3	
CIRCLE "1" ON PAGE F.3		CIRCLE "3" ON PAGE F.3	S9

A. MOOD EPISODES

PAST YEAR MAJOR DEPRESSIVE EPISODE

In the past year, that is since (CURRENT DATE) 2011, has there been a period of time when you were feeling depressed or down most of the day nearly every day? (What was that like?)

IF YES: When was that? How long did it last? (As long as two weeks?)

_____

_____

_____

_____

IF DEPRESSED MOOD: During that time did you lose interest or pleasure in things you usually enjoyed?

IF NO DEPRESSED MOOD: What about a time in the last year when you lost interest or pleasure in things you usually enjoyed? (What was that like?)

IF YES: When was that? Was it nearly every day? How long did it last? (As long as two weeks?)

Have you had more than one time in the past year like that? (Which time was the worst?)

_____

_____

_____

_____

_____

MDE CRITERIA

A. Five (or more) of the following symptoms have been present during the same two-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood, or (2) loss of interest or pleasure.

(1) depressed mood most of the day, nearly every day, as indicated either by subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). Note: in children or adolescents, can be irritable mood. ? 1 2 3 A1

(2) markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated either by subjective account or observation made by others). ? 1 2 3 A2

IF NEITHER ITEM (1) NOR ITEM (2) IS CODED "3," GO TO ***LIFETIME MDE,*** A.7

NOTE: IF MORE THAN ONE PAST EPISODE IS LIKELY, SELECT THE "WORST" ONE FOR YOUR INQUIRY ABOUT A MAJOR DEPRESSIVE EPISODE.

NOTE: WHEN RATING THE FOLLOWING ITEMS, CODE "1" IF CLEARLY DUE TO A GENERAL MEDICAL CONDITION, OR TO MOOD-INCONGRUENT DELUSIONS OR HALLUCINATIONS

FOR THE FOLLOWING QUESTIONS, FOCUS ON THE WORST TWO WEEKS OF THE MAJOR DEPRESSIVE EPISODE THAT YOU ARE INQUIRING ABOUT

During that (TWO WEEK PERIOD) . . .

. . . how was your appetite? (What about compared to your usual appetite?) (Did you have to force yourself to eat?) (Eat [less/more] than usual?) (Was that nearly every day?) (Did you lose or gain any weight) (How much?) (Were you trying to [lose/gain] weight?)	(3) significant weight loss when not dieting, or weight gain (e.g., a change of more than 5% of body weight in a month) or decrease or increase in appetite nearly every day. Note: in children, consider failure to make expected weight gains.	?	1	2	3	A3
. . . how were you sleeping? (Trouble falling asleep, waking frequently, trouble staying asleep, waking too early, OR sleeping too much? How many hours a night compared to usual? Was that nearly every night?)	(4) insomnia or hypersomnia nearly every day	?	1	2	3	A4
. . . were you so fidgety or restless that you were unable to sit still? (Was it so bad that other people noticed it? What did they notice? Was that nearly every day?) IF NO: What about the opposite -- talking or moving more slowly than is normal for you? (Was it so bad that other people noticed it? What did they notice? Was that nearly every day?)	(5) psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)	?	1	2	3	A5
. . . what was your energy like? (Tired all the time? Nearly every day?)	(6) fatigue or loss of energy nearly every day	?	1	2	3	A6

During this time . . .

. . .how did you feel about yourself? (Worthless?) (Nearly every day?) IF NO: What about feeling guilty about things you had done or not done? (Nearly every day?) _____ _____ _____ _____	(7) feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick) NOTE: CODE “1” OR “2” IF ONLY LOW SELF-ESTEEM	?	1	2	3	A7
. . .did you have trouble thinking or concentrating? (What kinds of things did it interfere with?) (Nearly every day?) IF NO: Was it hard to make decisions about everyday things? (Nearly every day?) _____ _____	(8) diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)	?	1	2	3	A8
. . .were things so bad that you were thinking a lot about death or that you would be better off dead? What about thinking of hurting yourself? IF YES: Did you do anything to hurt yourself? _____ _____ _____ _____ _____ _____	(9) recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide NOTE: CODE “1” FOR SELF-MUTILATION W/O SUICIDAL INTENT	?	1	2	3	A9

AT LEAST FIVE OF THE ABOVE SXS [A(1-9)] ARE CODED "3" AND AT LEAST ONE OF THESE IS ITEM (1) OR (2)

1 3 A10

IF NOT ALREADY ASKED: Has there been any other time in the past year when you were (depressed/OWN WORDS) and had even more of the symptoms that I just asked you about?

CONTINUE WITH NEXT ITEM, CRITERION C, MIDDLE OF PAGE

→ IF YES: RETURN TO * **PAST YEAR MDE*** A.1, AND CHECK WHETHER THERE HAVE BEEN ANY OTHER MAJOR DEPRESSIVE EPISODES IN THE PAST 12 MONTHS THAT WERE MORE SEVERE AND/OR CAUSED MORE SYMPTOMS. IF SO, ASK ABOUT THAT EPISODE.

→ IF NO: GO TO ***LIFETIME MDE*** A.7.

NOTE: DSM-IV criterion B (i.e., does not meet criteria for a mixed episode) has been omitted from the SCID.

C. The symptoms cause clinically significant distress or impairment in social, occupational or other important areas of functioning. 1 2 3 A11

IF UNCLEAR: Did (depressive episode/OWN WORDS) make it hard for you to do your work, take care of things at home, or get along with other people?

IF NOT ALREADY ASKED: Has there been any other time in the past year when you were (depressed/OWN WORDS) and it caused even more problems than the time I just asked you about?

CONTINUE ON NEXT PAGE

→ IF YES: RETURN TO ***PAST YEAR MDE*,** A.1, AND CHECK WHETHER THERE HAVE BEEN ANY OTHER MAJOR DEPRESSIVE EPISODES IN THE PAST 12 MONTHS THAT WERE MORE SEVERE AND/OR CAUSED MORE SYMPTOMS. IF SO, ASK ABOUT THAT EPISODE

→ IF NO: GO TO ***LIFETIME MDE*,** A.7.

In what month (and what year) did this (PAST YEAR MAJOR DEPRESSIVE EPISODE) start?

PAST YEAR MAJOR DEPRESSIVE EPISODE STARTED:

Month/Yr: ___ ___/___ ___ ___ ___ A12 A13

Just before this began, were you physically ill?

IF YES: What did the doctor say?

_____

Just before this began, were you using any medications?

IF YES: Was there any change in the amount you were taking at that time?

_____

Just before this began, were you drinking or using any street drugs?

_____

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition ? 1 3 A14

DUE TO SUB-STANCE USE OR GMC

IF THERE IS ANY INDICATION THAT THE DEPRESSION MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***MOOD EPISODES DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET**,** AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

PRIMARY MOOD EPISODE

IF UNKNOWN: Has there been any other time when you were (depressed / OWN WORDS) like this but were not (using SUBSTANCE / ill with GMC)?

→ IF YES: RETURN TO ***PAST YEAR MDE*,** A.1, AND CHECK WHETHER THERE HAVE BEEN ANY OTHER MAJOR DEPRESSIVE EPISODES IN THE PAST 12 MONTHS THAT WERE MORE SEVERE AND/OR CAUSED MORE SYMPTOMS. IF SO, ASK ABOUT THAT EPISODE

→ IF NO: GO TO ***LIFETIME MDE*,** A.7.

Etiological general medical conditions include: degenerative neurological illnesses (e.g., Parkinson's disease), cerebrovascular disease (e.g., stroke), metabolic conditions (e.g., Vitamin B-12 deficiency), endocrine conditions (e.g., hyper- and hypothyroidism, hyper- and hypoadrenocorticism); viral or other infections (e.g., hepatitis, mononucleosis, HIV), and certain cancers (e.g., carcinoma of the pancreas).

Etiological substances include: alcohol, amphetamines, cocaine, hallucinogens, inhalants, opioids, phencyclidine, sedatives, hypnotics, anxiolytics. Medications include antihypertensives, oral contraceptives, corticosteroids, anabolic steroids, anticancer agents, analgesics, anticholinergics, cardiac medications.

CONTINUE ON NEXT PAGE

Did this begin soon after someone close to you died?

E. The symptoms are not better accounted for by [Simple] Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms or psychomotor retardation.

NOTE: CODE "3" IF EITHER NOT FOLLOWING THE LOSS OF LOVED ONE OR IF BEREAVEMENT IS COMPLICATED BY MAJOR DEPRESSIVE EPISODE. CODE "1" IF SIMPLE BEREAVEMENT

1 → SIMPLE BEREAVE-MENT

3 → AT LEAST ONE EPISODE NOT SIMPLE BEREAVE-MENT

A15

IF UNKNOWN: Has there been any other time in the past year when you were (depressed / OWN WORDS) like this that did not occur after someone close to you died?

→ IF YES: GO TO ***PAST YEAR MDE *,** A.1 AND CHECK WHETHER THERE HAS BEEN ANY OTHER MAJOR DEPRESSIVE EPISODE IN THE PAST 12 MONTHS THAT WAS NOT BETTER ACCOUNTED FOR BY BEREAVEMENT. IF SO, ASK ABOUT THAT EPISODE.

→ IF NO: GO TO ***LIFETIME MDE*,** A.7.

CONTINUE BELOW

MAJOR DEPRESSIVE EPISODE CRITERIA A, C, D, AND E ARE CODED "3"

1 → GO TO ***LIFETIME MDE*,** A.7

3 → PAST YEAR MAJOR DEPRESSIVE EPISODE

A16

How many separate times in your life have you been (depressed / OWN WORDS) nearly every day for at least two weeks and had several of the symptoms that you described like (SXS OF WORST EPISODE)

Total number of Major Depressive Episodes (CODE 98 IF TOO NUMEROUS OR INDISTINCT TO COUNT) ____ ____

A17

GO TO ***PAST YEAR MANIC EPISODE*,** A.13

LIFETIME MAJOR DEPRESSIVE EPISODE

Looking back before the past year, have you *ever* had a period when you were feeling depressed or down most of the day nearly every day? (What was that like?)

IF YES: When was that? How long did it last? (As long as two weeks?)

_____

_____

IF PAST DEPRESSED MOOD: During that time, did you lose interest or pleasure in things you usually enjoyed? (What was that like?)

IF NO PAST DEPRESSED MOOD: Looking back before the past year, did you ever have a time when you lost interest or pleasure in things you usually enjoyed? (What was that like?)

IF YES: When was that? Was it nearly every day? How long did it last? (As long as two weeks?)

_____

_____

_____

_____

_____

Have you had more than one time like that? (Which time was the worst?)

MDE CRITERIA

A. Five or more of the following symptoms have been present during the same two-week period and represent a change from previous functioning; at least one of the symptoms was either (1) depressed mood or (2) loss of interest or pleasure.

? 1 2 3 A18

(1) depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad or empty) or observation made by others (e.g., appears tearful). Note: in children and adolescents, can be irritable mood.

? 1 2 3 A19

(2) markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated either by subjective account or observation made by others).

IF NEITHER ITEM (1) NOR (2) IS CODED "3", GO TO ***PAST YEAR MANIC EPISODE*** A.13

NOTE: IF MORE THAN ONE PAST EPISODE IS LIKELY, SELECT THE "WORST" ONE FOR YOUR INQUIRY ABOUT A MAJOR DEPRESSIVE EPISODE.

FOR THE FOLLOWING QUESTIONS, FOCUS ON THE WORST TWO WEEKS OF THE MAJOR DEPRESSIVE EPISODE THAT YOU ARE INQUIRING ABOUT

During that (TWO WEEK PERIOD) . . .

NOTE: WHEN RATING THE FOLLOWING ITEMS, CODE "1" IF CLEARLY DIRECTLY DUE TO A GENERAL MEDICAL CONDITION, OR TO MOOD-INCONGRUENT DELUSIONS OR HALLUCINATIONS

. . . how was your appetite? (What about compared to your usual appetite?) (Did you have to force yourself to eat?) (Eat [less/more] than usual?) (Was that nearly every day?) (Did you lose or gain any weight?) (How much?) (Were you trying to [lose/gain] weight?) _____ _____ _____	(3)	significant weight loss when not dieting, or weight gain (e.g., a change of more than 5% of body weight in a month) or decrease or increase in appetite nearly every day.	?	1	2	3	A20
. . .how were you sleeping? (Trouble falling asleep, waking frequently, trouble staying asleep, waking too early, OR sleeping too much? How many hours a night compared to usual? Was that nearly every night? _____ _____	(4)	insomnia or hypersomnia nearly every day	?	1	2	3	A21
. . . were you so fidgety or restless that you were unable to sit still? (Was it so bad that other people noticed it? What did they notice? Was that nearly every day?) IF NO: What about the opposite -- talking or moving more slowly than is normal for you? (Was it so bad that other people noticed it? What did they notice? Was it nearly every day?)	(5)	psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down)	?	1	2	3	A22
. . . what was your energy like? (Tired all the time? Nearly every day?) _____	(6)	fatigue or loss of energy nearly every day	?	1	2	3	A23

During that time . . .

. . . how did you feel about yourself? (Worthless?) (Nearly every day?) _____ _____ _____ IF NO: What about feeling guilty about things you had done or not done? (Nearly every day?) _____ _____	(7) feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick) NOTE: CODE "1" OR "2" FOR LOW SELF-ESTEEM BUT NOT WORTHLESSNESS	?	1	2	3	A24
. . . did you have trouble thinking or concentrating? (What kinds of things did it interfere with?) (Nearly every day?) _____ _____ _____ IF NO: Was it hard to make decisions about everyday things? (Nearly every day?) _____ _____	(8) diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others)	?	1	2	3	A25
. . .were things so bad that you were thinking a lot about death or that you would be better off dead? What about thinking of hurting yourself? _____ _____ _____ IF YES: Did you do anything to hurt yourself? _____ _____	(9) recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide NOTE: CODE "1" FOR SELF-MUTILATION W/O SUICIDAL INTENT	?	1	2	3	A26

AT LEAST FIVE OF THE ABOVE SXS [A(1-9)] ARE CODED "3" AND AT LEAST ONE OF THESE IS ITEM (1) OR (2)

1 3 A27

IF NOT ALREADY ASKED: Has there been any other time when you were (depressed/OWN WORDS) and had even more of the symptoms that I just asked you about?

CONTINUE WITH NEXT ITEM, CRITERION C, MIDDLE OF PAGE

IF YES: RETURN TO ***LIFETIME MDE*,** A.7, AND CHECK WHETHER THERE HAVE BEEN ANY OTHER MAJOR DEPRESSIVE EPISODES THAT WERE MORE SEVERE AND/OR CAUSED MORE SYMPTOMS. IF SO, ASK ABOUT THAT EPISODE.

IF NO: GO TO ***PAST YEAR MANIC EPISODE*,** A.13.

NOTE: DSM-IV criterion B (i.e., does not meet criteria for a mixed episode) has been omitted from the SCID.

IF UNCLEAR: Did (depressive episode/OWN WORDS) make it hard for you to do your work, take care of things at home, or get along with other people?

C. The symptoms cause clinically significant distress or impairment in social, occupational or other important areas of functioning. 1 2 3 A28

IF NOT ALREADY ASKED: Has there been any other time when you were (depressed/OWN WORDS) and it caused even more problems than the time I just asked you about?

CONTINUE ON NEXT PAGE

IF YES: RETURN TO ***LIFETIME MDE*,** A.7, AND CHECK WHETHER THERE HAVE BEEN ANY OTHER MAJOR DEPRESSIVE EPISODES THAT WERE MORE SEVERE AND/OR CAUSED MORE SYMPTOMS. IF SO, ASK ABOUT THAT EPISODE

IF NO: GO TO ***PAST YEAR MANIC EPISODE*,** A.13.

How old were you when this (LIFETIME MAJOR DEPRESSIVE EPISODE) started?

LIFETIME MAJOR DEPRESSIVE EPISODE STARTED:

AGE: ___ ___ A29

Just before this began, were you physically ill?

IF YES: What did the doctor say

_____

Just before this began, were you using any medications?

IF YES: Was there any change in the amount you were taking at that time?

_____

Just before this began, were you drinking or using any street drugs?

_____

D. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition (e.g., hypothyroidism) ? 1 3 A30

DUE TO SUBSTANCE USE OR GMC

IF THERE IS ANY INDICATION THAT THE DEPRESSION MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***MOOD EPISODE DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

REFER TO LIST OF GENERAL MEDICAL CONDITIONS AND SUBSTANCES, A.5.

PRIMARY MOOD EPISODE

IF UNKNOWN: Has there been any other time when you were (depressed/ OWN WORDS) like this but were not (using SUBSTANCE / ill with GMC)?

→ IF YES: GO TO ***LIFETIME MDE*,** A.7 AND CHECK WHETHER THERE HAS BEEN ANY OTHER MAJOR DEPRESSIVE EPISODE NOT DUE TO A SUBSTANCE OR GENERAL MEDICAL CONDITION. IF SO, ASK ABOUT THAT EPISODE.

→ IF NO: GO TO ***PAST YEAR MANIC EPISODE*,** A.13

CONTINUE ON NEXT PAGE

Did this begin soon after someone close to you died?

_____

_____

E. The symptoms are not better accounted for by [Simple] Bereavement, i.e., after the loss of a loved one, the symptoms persist for longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms or psychomotor retardation.

NOTE: CODE "3" IF EITHER NOT FOLLOWING THE LOSS OF LOVED ONE OR IF BEREAVEMENT IS COMPLICATED BY MAJOR DEPRESSIVE EPISODE. CODE "1" IF SIMPLE BEREAVEMENT

1 → SIMPLE BEREAVEMENT

3 → AT LEAST ONE EPISODE NOT SIMPLE BEREAVEMENT → CONTINUE BELOW

A31

IF UNKNOWN: Has there been any other time when you were (depressed/ OWN WORDS) like this that did not occur after someone close to you died?

→ IF YES: GO TO ***LIFETIME MDE*,** A.7 AND CHECK WHETHER THERE HAS BEEN ANY OTHER MAJOR DEPRESSIVE EPISODE THAT WAS NOT BETTER ACCOUNTED FOR BY BEREAVEMENT. IF SO, ASK ABOUT THAT EPISODE.

→ IF NO: GO TO ***PAST YEAR MANIC EPISODE*,** A.13.

MAJOR DEPRESSIVE EPISODE CRITERIA A, C, D, AND E ARE CODED "3"

1 → GO TO ***PAST YEAR MANIC EPISODE*,** A.13

3 → LIFETIME MAJOR DEPRESSIVE EPISODE

A32

How many separate times in your life have you been (depressed/OWN WORDS) nearly every day for at least two weeks and had several of the symptoms that you described like (SXS OF WORST EPISODE)

Total number of Major Depressive Episodes (CODE 98 IF TOO NUMEROUS OR INDISTINCT TO COUNT)

____ ____

→ GO TO ***PAST YEAR MANIC EPISODE*,** A.13

A34

PAST YEAR MANIC EPISODE

In the past year has there been a period of time when you were feeling so good, "high," excited, or hyper that other people thought you were not your normal self or you were so hyper that you got into trouble?

IF YES: What was it like? (Did anyone say you were manic?) (Was that more than just feeling good?)

IF NO: In the past year, have you had a period of time when you were feeling irritable or angry every day for at least several days?

What was it like? (Did you find yourself often starting fights or arguments?)

_____

_____

_____

_____

How long did that last? (As long as one week?) (Did you have to go into a hospital?)

_____

_____

_____

_____

_____

_____

MANIC EPISODE CRITERIA

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood — ? 1 2 3 — A35

GO TO ***LIFETIME MANIC EPISODE*,** A.19

. . . lasting at least one week (or any duration if hospitalization is necessary) — ? 1 2 3 — A36

NOTE: IF ELEVATED MOOD LASTS LESS THAN ONE WEEK, CHECK WHETHER IRRITABLE MOOD LASTS AT LEAST ONE WEEK BEFORE SKIPPING TO A.19.

GO TO ***LIFETIME MANIC EPISODE*,** A.19

FOCUS ON THE WORST PERIOD OF THE EPISODE THAT YOU ARE INQUIRING ABOUT.

IF UNCLEAR: During (EPISODE), when were you the most (OWN WORDS FOR MANIA)?

During that time . . .

B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

. . .how did you feel about yourself? (More self-confident than usual?) (Any special powers or abilities?) _____ _____	(1)	inflated self-esteem or grandiosity	?	1	2	3	A37
. . .did you need less sleep than usual? (How much sleep did you get?) IF YES: Did you still feel rested? _____ _____	(2)	decreased need for sleep (e.g., feels rested after only three hours of sleep)	?	1	2	3	A38
. . .were you much more talkative than usual? (Did people have trouble stopping you or understanding you? Did people have trouble getting a word in edgewise?) _____ _____	(3)	more talkative than usual or pressure to keep talking	?	1	2	3	A39
. . .were your thoughts racing through your head? (What was that like?) _____ _____	(4)	flight of ideas or subjective experience that thoughts are racing	?	1	2	3	A40

During that time . . .

. . .were you so easily distracted by things around you that you had trouble concentrating or staying on one track? (Give me an example of that.) _____ _____	(5)	distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)	?	1	2	3	A41
. . .how did you spend your time? (Work, friends, hobbies?) (Were you especially productive or busy during that time?) (Were you so active that your friends or family were concerned about you?) IF NO INCREASED ACTIVITY: Were you physically restless? (How bad was it?) _____ _____	(6)	increase in goal- directed activity (either socially, at work or school, or sexually) or psychomotor agitation	?	1	2	3	A42
. . .did you do anything that could have caused trouble for you or your family? (Buying things you didn't need?) (Anything sexual that was unusual for you?) (Reckless driving?) _____ _____	(7)	excessive involvement in pleasurable activities which have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments)	?	1	2	3	A43

AT LEAST THREE "B" SXS ARE CODED "3" (FOUR IF MOOD ONLY IRRITABLE) 1 3 A44

IF NOT ALREADY ASKED: Has there been any other times in the past year when you were (high/irritable/OWN WORDS) and had even more of the symptoms that I just asked you about? A44a

→ IF YES: RETURN TO ***PAST YEAR MANIC EPISODE*,** A.13, AND INQUIRE ABOUT WORST EPISODE.

→ IF NO: GO TO ***LIFETIME MANIC EPISODE*,** A.19.

NOTE: DSM-IV criterion C (i.e., does not meet criteria for a Mixed Episode) has been omitted from the SCID.

CONTINUE BELOW

IF NOT KNOWN: At that time, did you have serious problems at home or at work (school) because you were (SYMPTOMS) or did you have to go into a hospital?

D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others or there are psychotic features. ? 1 2 3 A45

IF NOT ALREADY ASKED: Have there been any other times in the past year when you were (high/irritable/OWN WORDS) and had (ACKNOWLEDGED MANIC SYMPTOMS) and you got into trouble with people or were hospitalized?

DESCRIBE:

→ IF YES: RETURN TO ***PAST YEAR MANIC EPISODE*,** A.13 AND INQUIRE ABOUT THAT EPISODE

→ IF NO: GO TO ***LIFETIME MANIC EPISODE*,** A.19.

CONTINUE ON NEXT PAGE

In what month (and what year) did this (PAST YEAR MANIC EPISODE) start?

PAST YEAR MANIC EPISODE STARTED:

Month/Yr: ___ ___/___ ___ ___ ___ A46 A47

Just before this began, were you physically ill?

IF YES: What did the doctor say?

Just before this began, were you taking any medications?

IF YES: Was there any change in the amount you were taking at that time?

Just before this began, were you drinking or using any street drugs?

NOTE: MANIC-LIKE EPISODES THAT ARE CLEARLY CAUSED BY SOMATIC ANTIDEPRESSANT TREATMENT (E.G., MEDICATION, ECT, LIGHT THERAPY) SHOULD NOT COUNT TOWARD A DIAGNOSIS OF BIPOLAR I DISORDER BUT ARE CONSIDERED SUBSTANCE-INDUCED MOOD DISORDERS.

E. The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition — ? 1 3 A48

DUE TO SUBSTANCE USE OR GMC

IF THERE IS ANY INDICATION THAT THE MANIA MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***MOOD EPISODE DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

Etiological general medical conditions include: degenerative neurological illnesses (e.g., Huntington's disease, multiple sclerosis), cerebrovascular disease (e.g., stroke), metabolic conditions (e.g., Vitamin B-12 deficiency, Wilson's disease), endocrine conditions (e.g., hyperthyroidism), viral or other infections, and certain cancers (e.g., cerebral neoplasms).

IF UNKNOWN: Has there been any other times in the past year when you were (high / irritable / OWN WORDS) and were not (using SUBSTANCE / ill with GMC)?

Etiological substances include: alcohol, amphetamines, cocaine, hallucinogens, inhalants, opioids, phencyclidine, sedatives, hypnotics, and anxiolytics. Medications include psychotropic medications (e.g., anti-depressants), corticosteroids, anabolic steroids, isoniazid, antiparkinson medication (e.g., levadopa), and sympathomimetics/decongestants

PRIMARY MOOD EPISODE

→ IF YES: RETURN TO ***PAST YEAR MANIC EPISODE*,** A.13, AND INQUIRE ABOUT OTHER EPISODE.

→ IF NO: GO TO ***LIFETIME MANIC EPISODE*,** A.19.

MANIC EPISODE CRITERIA A, B, D AND E ARE CODED "3" — 1 3 A49

GO TO ***LIFETIME MANIC EPISODE*,** A.19

MANIC EPISODE PAST YEAR

How many separate times in your life were you (HIGH/OWN WORDS) and had [ACKNOWLEDGED MANIC SYMPTOMS] for at least a week (or were hospitalized)?	Number of Manic Episodes, including past year (CODE 98 IF TOO INDISTINCT OR NUMEROUS TO COUNT)	____ ____	A50

GO TO ***PSYCHOTIC SCREEN*,** B/C.1

LIFETIME MANIC EPISODE

Looking back before the past year, did you *ever* have a period of time when you were feeling so good, "high," excited, or hyper that other people thought you were not your normal self or you were so hyper that you got into trouble?

IF YES: What was it like? (Did anyone say you were manic?) (Was that more than just feeling good?)

IF NO: Looking back before the past year, did you *ever* have a period of time when you were feeling irritable or angry every day for at least several days?

What was it like? (Did you find yourself often starting fights or arguments?)

_____

_____

When was that?

_____

How long did that last? (as long as one week?) (Did you need to go to the hospital?)

_____

_____

Have you had more than one time like that? (Which time was the most extreme?)

_____

_____

MANIC EPISODE CRITERIA

A. A distinct period of abnormally and persistently elevated, expansive, or irritable mood . . . ? 1 2 3 A51

GO TO ***DYSTHYMIC DISORDER*,** A.25

. . . lasting at least one week (or any duration if hospitalization is necessary) ? 1 2 3 A52

GO TO ***DYSTHYMIC DISORDER*,** A.25

NOTE: IF ELEVATED MOOD LASTS LESS THAN ONE WEEK, CHECK WHETHER IRRITABLE MOOD LASTS AT LEAST ONE WEEK BEFORE SKIPPING TO A.25.

NOTE: IF THERE IS EVIDENCE FOR MORE THAN ONE PAST EPISODE, SELECT THE "WORST" ONE FOR YOUR INQUIRY ABOUT PAST MANIC EPISODE.

FOCUS ON THE WORST PERIOD OF THE EPISODE THAT YOU ARE INQUIRING ABOUT.

IF UNCLEAR: During (EPISODE), when were you the most (OWN WORDS FOR MANIA)?

During that time . . .

B. During the period of mood disturbance, three (or more) of the following symptoms have persisted (four if the mood is only irritable) and have been present to a significant degree:

. . .how did you feel about yourself? (More self-confident than usual?) (Any special powers or abilities?)	(1)	inflated self-esteem or grandiosity	?	1	2	3	A53
. . .did you need less sleep than usual? (How much sleep did you get?) IF YES: Did you still feel rested?	(2)	decreased need for sleep (e.g., feels rested after only three hours of sleep)	?	1	2	3	A54
. . .were you much more talkative than usual? (Did people have trouble stopping you or understanding you? Did people have trouble getting a word in edgewise?)	(3)	more talkative than usual or pressure to keep talking	?	1	2	3	A55
. . .were your thoughts racing through your head? (What was that like?)	(4)	flight of ideas or subjective experience that thoughts are racing	?	1	2	3	A56
. . .were you so easily distracted by things around you that you had trouble concentrating or staying on one track? (Give me an example of that.)	(5)	distractibility (i.e., attention too easily drawn to unimportant or irrelevant external stimuli)	?	1	2	3	A57
. . .how did you spend your time? (Work, friends, hobbies?) (Were you especially productive or busy during that time?) (Were you so active that your friends or family were concerned about you?) IF NO INCREASED ACTIVITY: Were you physically restless? (How bad was it?)	(6)	increase in goal- directed activity (either socially, at work or school, or sexually) or psychomotor agitation	?	1	2	3	A58

During that time . . .

. . .did you do anything that could have caused trouble for you or your family? (Buying things you didn't need?) (Anything sexual that was unusual for you?) (Reckless driving?)

_____

_____

(7) excessive involvement in pleasurable activities which have a high potential for painful consequences (e.g., engaging in unrestrained buying sprees, sexual indiscretions, or foolish business investments) ? 1 2 3 A59

AT LEAST THREE "B" SXS ARE CODED "3" (FOUR IF MOOD ONLY IRRITABLE) 1 3 A60

IF NOT ALREADY ASKED: Has there been any other time when you were (high/irritable/OWN WORDS) and had even more of the symptoms that I just asked you about? A60a

→ IF YES: RETURN TO ***LIFETIME MANIC EPISODE*,** A.19, AND INQUIRE ABOUT WORST EPISODE.

→ IF NO: GO TO ***DYSTHYMIC DISORDER*,** A.25.

NOTE: DSM-IV criterion C (i.e., does not meet criteria for a Mixed Episode) has been omitted from the SCID.

CONTINUE BELOW

IF NOT KNOWN: At that time, did you have serious problems at home or at work (school) because you were (SYMPTOMS) or did you have to go into a hospital?

_____

D. The mood disturbance is sufficiently severe to cause marked impairment in occupational functioning or in usual social activities or relationships with others, or to necessitate hospitalization to prevent harm to self or others or there are psychotic features. 1 2 3 A61

IF NOT ALREADY ASKED: Has there been any other time when you were (high/irritable/OWN WORDS) and had (ACKNOWLEDGED MANIC SYMPTOMS) and you got into trouble with people or were hospitalized?

DESCRIBE:

→ IF YES: RETURN TO ***LIFETIME MANIC EPISODE",** A.19 AND INQUIRE ABOUT THAT EPISODE

→ IF NO: GO TO ***DYSTHYMIC DISORDER*,** A.25.

CONTINUE ON NEXT PAGE

How old were you when this (LIFETIME MANIC EPISODE) started?

LIFETIME MANIC EPISODE STARTED:

AGE: ___ ___ A62

Just before this began, were you physically ill?

IF YES: What did the doctor say?

_____

Just before this began, were you taking any medications?

IF YES: Was there any change in the amount you were taking at that time?

_____

Just before this began, were you drinking or using any street drugs?

_____

_____

_____

_____

E The symptoms are not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition ? 1 3 A63

DUE TO SUBSTANCE USE OR GMC

IF THERE IS ANY INDICATION THAT THE MANIA MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***MOOD EPISODE DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

NOTE: MANIC-LIKE EPISODES THAT ARE CLEARLY CAUSED BY SOMATIC ANTIDEPRESSANT TREATMENT (E.G., MEDICATION, ECT, LIGHT THERAPY) SHOULD NOT COUNT TOWARD A DIAGNOSIS OF BIPOLAR I DISORDER BUT ARE CONSIDERED SUBSTANCE-INDUCED MOOD DISORDERS.

REFER TO LIST OF GENERAL MEDICAL CONDITIONS AND SUBSTANCES, A.17

PRIMARY MOOD EPISODE

IF UNKNOWN: Has there been any other time when you were (high / irritable / OWN WORDS) and were not (using SUBSTANCE / ill with GMC)?

→ IF YES: RETURN TO ***LIFETIME MANIC EPISODE*,** A.19, AND INQUIRE ABOUT OTHER EPISODE.

→ IF NO: GO TO ***DYSTHYMIC DISORDER*,** A.25.

CONTINUE ON NEXT PAGE

MANIC EPISODE CRITERIA
A, B, D AND E ARE CODED "3"

1 — GO TO ***DYSTHYMIC DISORDER***, A.25

3 — LIFETIME MANIC EPISODE

A64

How many separate times in your life were you (HIGH / OWN WORDS) and had [ACKNOWLEDGED MANIC SYMPTOMS] for a period of time (or were hospitalized)?

Number of Manic Episodes (CODE 98 IF TOO INDISTINCT OR NUMEROUS TO COUNT)

____ ____ A65

GO TO ***PSYCHOTIC SCREEN***, B/C.1

This page has been intentionally left blank.

DYSTHYMIC DISORDER
(PAST YEAR)

DYSTHYMIC DISORDER CRITERIA

For the past couple of years, have you been bothered by depressed mood most of the day, more days than not? (More than half of the time?)

IF YES: What was that like?

_____

A. Depressed mood for most of the day, for more days than not, as indicated either by subjective account or observation made by others, for at least two years. Note: in children and adolescents, mood can be irritable and duration must be at least 1 year.

? 1 2 3 A66

GO TO ***PSYCHOTIC SCREEN***, B/C.1

During these periods of (OWN WORDS FOR CHRONIC DEPRESSION) do you often . . .

B Presence, while depressed, of two (or more) of the following:

. . . lose your appetite? (What about overeating?)

_____

_____

(1) poor appetite or overeating ? 1 2 3 A67

. . . have trouble sleeping or sleep too much?

_____

_____

(2) insomnia or hypersomnia ? 1 2 3 A68

. . . have little energy to do things or feel tired a lot?

_____

_____

(3) low energy or fatigue ? 1 2 3 A69

. . . feel down on yourself? (Feel worthless, or a failure?)

_____

_____

(4) low self-esteem ? 1 2 3 A70

. . . have trouble concentrating or making decisions?

_____

_____

(5) poor concentration or difficulty making decisions ? 1 2 3 A71

. . . feel hopeless? _____ _____	(6) feelings of hopelessness	? 1 2 3	A72
	AT LEAST TWO "B" SYMPTOMS CODED "3"	? 1 2 3 GO TO ***PSYCHOTIC SCREEN***, B/C.1	A73
What is the longest period of time, during this period of long-lasting depression, that you felt OK? (NO DYSTHYMIC SYMPTOMS)	C. During the two year period of the disturbance, the person has never been without the symptoms in criteria A and B for more than two months at a time. NOTE: CODE "1" IF NORMAL MOOD FOR AT LEAST TWO MONTHS AT A TIME	1 3 GO TO ***PSYCHOTIC SCREEN***, B/C.1	A74
How long have you been feeling this way? (When did this begin?) COMPARE ONSET OF DYSTHYMIC SXS WITH DATES OF PAST MAJOR DEPRESSIVE EPISODES TO DETERMINE IF THERE WERE ANY MAJOR DEPRESSIVE EPISODES IN FIRST TWO YEARS OF DYSTHYMIC DISORDER.	D. No Major Depressive Episode has been present during the first 2 years of the disturbance (1 year for children and adolescents): i.e., not better accounted for by chronic Major depressive Disorder or Major Depression in partial remission.	1 3 GO TO ***PSYCHOTIC SCREEN***, B/C.1	A75
	Age at onset of current Dysthymic Disorder (CODE 98 IF UNKNOWN)	____ ____	A76
IF A MAJOR DEPRESSIVE EPISODE PRECEDED DYSTHYMIC SXS: Now I want to know whether you got completely back to your usual self after that (MAJOR DEPRESSIVE EPISODE) you had (DATE), before this long period of being mildly depressed? (Were you back to your usual self for at least two months?)	Note: There may have been a previous Major Depressive Episode provided there was a full remission (no significant signs or symptoms for 2 months) before development of the Dysthymic Disorder. In addition, after the initial 2 years (1 year for children or adolescents) of Dysthymic Disorder, there may be superimposed episodes of Major Depressive Disorder, in which case both diagnoses may be given when criteria are met for a Major Depressive Episode.		

NOTE: CODE “3” IF NO PAST MAJOR DEPRESSIVE EPISODES OR IF MAJOR DEPRESSIVE EPISODES WERE NOT PRESENT DURING THE FIRST TWO YEARS OR IF THERE WAS AT LEAST A TWO-MONTH PERIOD WITHOUT SYMPTOMS PRECEDING THE ONSET.

E. NOTE: RULE OUT FOR MIXED EPISPODE AND HYPOMANIC EPISODE AND CYCLOTHYMIC DISORDER ARE NOT ASSESSED HERE.

IF NOT ALREADY CLEAR: RETURN TO THIS ITEM AFTER COMPLETING THE PSYCHOTIC SYMPTOMS SECTION.

F. The disturbance does not occur exclusively during the course of a chronic psychotic disorder, such as Schizophrenia or Delusional Disorder.

NOTE: CODE “3” IF NO CHRONIC PSYCHOTIC DISORDER OR IF NOT SUPERIMPOSED ON A CHRONIC PSYCHOTIC DISORDER

1	3	
GO TO *PSYCHOTIC SCREEN*, B/C.1	NOT SUPER-IMPOSED CON-TINUE	A77

Just before this began, were you physically ill?

IF YES: What did the doctor say?

Just before this began, were you using any medications?

IF YES: Was there any change in the amount you were taking at that time?

_____

_____

Just before this began, were you drinking or using any street drugs?

G. Not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition

? 1 3 A78

DUE TO SUBSTANCE USE OR GMC

GO TO ***PSYCHOTIC SCREEN***, B/C.1

IF THERE IS ANY INDICATION THAT THE DYSTHMIA MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***MOOD EPISODE DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

PRIMARY MOOD DISORDER

Etiological general medical conditions include: degenerative neurological illnesses (e.g., Parkinson's disease, Huntington's disease, cerebrovascular disease, metabolic and endocrine conditions (e.g., B-12 deficiency, hypothyroidism, autoimmune conditions (e.g., systemic lupus erythematosis), viral or other infections (e.g., hepatitis, mononucleosis, HIV), and certain cancers (e.g., carcinoma of the pancreas)

Etiological substances include: alcohol, amphetamines, cocaine, hallucinogens, inhalants, opioids, phencyclidine, sedatives, hypnotics, anxiolytics. Medications include antihypertensives, oral contraceptives, corticosteroids, anabolic steroids, anticancer agents, analgesics, anti-cholinergics, and cardiac medications.

CONTINUE BELOW

IF UNCLEAR: How much do your depressed feelings interfere with your life?

H. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning

? 1 2 3 A79

GO TO ***PSYCHOTIC SCREEN***, B/C.1

DYSTHYMIC DISORDER CRITERIA A, B, C, D, F, G, AND H ARE CODED "3."

1 3 A80

GO TO ***PSYCHOTIC SCREEN***, B/C.1

DYSTHYMIC DISORDER

B/C PSYCHOTIC SCREENING MODULE

THIS MODULE IS FOR CODING PSYCHOTIC AND ASSOCIATED SXS THAT HAVE BEEN PRESENT AT ANY POINT IN THE PAST YEAR.

FOR EACH PSYCHOTIC SYMPTOM CODED "3," DESCRIBE THE ACTUAL CONTENT AND INDICATE THE PERIOD OF TIME DURING WHICH THE SYMPTOM WAS PRESENT.

FOR ANY DELUSIONS OR HALLUCINATIONS CODED “3”, DETERMINE WHETHER THE SYMPTOM IS DEFINITELY “PRIMARY” OR WHETHER THERE IS A POSSIBLE OR DEFINITE ETIOLOGIC SUBSTANCE (INCLUDING MEDICATIONS) OR GENERAL MEDICAL CONDITION. THE FOLLOWING QUESTIONS MAY BE USEFUL IF THE OVERVIEW HAS NOT ALREADY PROVIDED THE INFORMATION:

Just before (PSYCHOTIC SXS) began, were you using drugs? ...on any medications? ...did you drink much more than usual or stop drinking after you had been drinking a lot for a while? ...were you physically ill?

IF YES TO ANY: Has there been a time when you had (PSYCHOTIC SXS) and were not (USING DRUGS/TAKING MEDICATION/CHANGING YOUR DRINKING HABITS/ILL)?

Now I am going to ask you about unusual experiences that people sometimes have.

In the past year, that is since (CURRENT DATE) 2011…

…did it ever seem like people were talking about you or taking special notice of you?

IF YES: Were you convinced they were talking about you or did you think it might have been your imagination?

…what about receiving special messages from the TV, radio, or newspaper, or from the way things were arranged around you?
DESCRIBE:

_____

_____

_____

_____

DELUSIONS
False personal beliefs based on incorrect inference about external reality and firmly sustained in spite of what almost everyone else believes and in spite of what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture. Code overvalued ideas (unreasonable and sustained beliefs that are maintained with less than delusional intensity) as "2."

Delusion of reference, i.e., events, objects, or other people in the individual's immediate environment have a particular or unusual significance.

? 1 2 3 BC1

1	3
POSS/DEF SUBST/ GMC	PRI-MARY

BC2

In the past year…

…have you felt that someone was going out of their way to give you a hard time, or trying to hurt you?

DESCRIBE:

_____

_____

_____

Persecutory delusion, i.e., the individual (or his or her group) is being attacked, harassed, cheated, persecuted, or conspired against.

? 1 2 3 BC3

1 3 BC4

POSS/DEF SUBST/ GMC PRI-MARY

…have you felt that you were especially important in some way, or that you had special powers to do things that other people couldn't do?

DESCRIBE:

_____

_____

_____

Grandiose delusion, i.e., content involves exaggerated power, knowledge or importance, or a special relationship to a deity or famous person.

? 1 2 3 BC5

1 3 BC6

POSS/DEF SUBST/ GMC PRI-MARY

In the past year have you felt that something was very wrong with you physically even though your doctor said nothing was wrong...like you had cancer or some other terrible disease?

…have you been convinced that something was very wrong with the way a part or parts of your body looked?

…have you felt that something strange was happening to parts of your body?

DESCRIBE:

_____

_____

_____

Somatic delusion, i.e., content involves change or disturbance in body appearance or functioning.

? 1 2 3 BC7

1 3 BC8

POSS/DEF SUBST/ GMC PRI-MARY

In the past year…

…have you had any unusual religious experiences?

…have you felt that you had committed a crime or done something terrible for which you should be punished?

…have you been convinced that your spouse or partner was being unfaithful to you?

IF YES: How did you know they were being unfaithful?

…did you feel you had a special, secret relationship with someone famous, or someone you didn't know very well?

DESCRIBE:

_____

_____

_____

Other delusions

? 1 2 3 BC9

1	3
POSS/DEF SUBST/ GMC	PRI-MARY

BC10

HALLUCINATIONS (PSYCHOTIC):
A sensory perception that has the compelling sense of reality of a true perception but occurs without external stimulation of the relevant sensory organ. (CODE "2" FOR HALLUCINATIONS THAT ARE SO TRANSIENT AS TO BE WITHOUT DIAGNOSTIC SIGNIFICANCE)

In the past year…

…have you heard things that other people couldn't hear, such as noises, or the voices of people whispering or talking? (Were you awake at the time?)

IF YES: What did you hear? How often did you hear it?

DESCRIBE:

_____

_____

_____

Auditory hallucinations when fully awake, heard either inside or outside of head — ? 1 2 3 — BC11

1 POSS/DEF SUBST/ GMC — 3 PRI-MARY — BC12

IF VOICES: Did they comment on what you were doing or thinking?

A voice keeping up a running commentary on the individual's behavior or thoughts as they occur — ? 1 2 3 — BC13

How many voices did you hear? Were they talking to each other?

Two or more voices conversing with each other — ? 1 2 3 — BC14

How about having visions or seeing things that other people couldn't see? (Were you awake at the time?)

NOTE: DISTINGUISH FROM AN ILLUSION, I.E., A MISPERCEPTION OF A REAL EXTERNAL STIMULUS.

DESCRIBE:

_____

_____

_____

Visual hallucinations — ? 1 2 3 — BC15

1 POSS/DEF SUBST/ GMC — 3 PRI-MARY — BC16

…what about strange sensations in your body or on your skin?

DESCRIBE:

_____

_____

_____

Tactile hallucinations, e.g., electricity

? 1 2 3 BC17

1 3 BC18

POSS/DEF SUBST/ GMC | PRI-MARY

(What about smelling or tasting things that other people couldn't smell or taste?)

DESCRIBE:

_____

_____

_____

Other hallucinations, e.g., gustatory, olfactory

? 1 2 3 BC19

1 3 BC20

POSS/DEF SUBST/ GMC | PRI-MARY

ANY ITEM CODED "3" IN "PRIMARY" SECTION

? 1 3 BC21

GO TO ***MOOD DISORDERS***, D.1

A PRI-MARY PSYCHO-TIC SX HAS BEEN PRESENT

EXPLORE DETAILS AND DESCRIBE DIAGNOSTIC SIGNIFICANCE:

_____

_____

_____

This page has been intentionally left blank.

D. MOOD DISORDERS

IF THERE HAVE NEVER BEEN ANY CLINICALLY SIGNIFICANT MOOD SYMPTOMS, CIRCLE 1 AND GO TO ***PTSD,*** E.1. 1 3 D1

IF THERE HAVE BEEN ANY CLINICALLY SIGNIFICANT MOOD SYMPTOMS, CIRCLE 3 AND CONTINUE.

BIPOLAR I DISORDER CRITERIA

CODE BASED ON RATINGS OF MANIC EPISODE PAST YEAR (A49) AND MANIC EPISODE LIFETIME (A64)

History of one or more Manic or Mixed Episodes 1 3 D2

GO TO ***MAJOR DEPRESSIVE DISORDER*,** D. 2

Note: In a Mixed Episode, the criteria are met for both a Manic Episode and a Major Depressive Episode (except for duration nearly every day during at least a 1-week period

At least one Manic or Mixed Episode is not due to the direct physiological effects of a general medical condition or substance use 1 3 D3

GO TO ***MAJOR DEPRESSIVE DISORDER*,** D.2

Note: Manic-like Episodes that are clearly caused by somatic antidepressant treatment (e.g., medication, ECT, light therapy) should not count toward a diagnosis of Bipolar I Disorder

At least one Manic or Mixed Episode is not better accounted for by Schizoaffective Disorder and is not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder NOS 1 3 D4

GO TO ***MAJOR DEPRESSIVE DISORDER*,** D. 2

BIPOLAR I DISORDER

Indicate time frame of manic episode: D5
1 – Manic episode in past 12 months
2 – Manic episode lifetime (i.e., prior to past 12 months)

GO TO ***PTSD*,** E.1

MAJOR DEPRESSIVE DISORDER

CODE BASED ON RATINGS OF PAST YEAR MAJOR DEPRESSIVE EPISODE (A16) AND LIFETIME MAJOR DEPRESSIVE DISODER (A32)

MAJOR DEPRESSIVE DISORDER CRITERIA

At least one Major Depressive Episode that is <u>not</u> due to the direct physiological effects of a general medical condition or substance use	1 → GO TO ***PTSD***, E.1	3	D6
At least one Major Depressive Episode that is <u>not</u> better accounted for by Schizoaffective Disorder and is not superimposed on Schizophrenia, Schizophreniform Disorder, Delusional Disorder, or Psychotic Disorder Not Otherwise Specified	1 → GO TO ***PTSD***, E.1	3	D7
Has never had any Manic, Mixed, or unequivocal Hypomanic Episodes	1 → GO TO ***PTSD***, E.1	3 → MAJOR DEPRESSIVE DISORDER	D8

Indicate type:

1 - Single Episode

2 - Recurrent (i.e., to be considered separate episodes, there must be an interval of at least two months in which criteria are not met for a Major Depressive Episode)

D9

GO TO ***PTSD*, E.1**

E. ANXIETY DISORDERS

POSTTRAUMATIC STRESS DISORDER

Sometimes things happen to people that are extremely upsetting--things like being in a life threatening situation like a major disaster, very serious accident or fire; being physically assaulted or raped; seeing another person killed or dead, or badly hurt, or hearing about something horrible that has happened to someone you are close to. At any time during your life, have any of these kinds of things happened to you?

IF NO: Have you ever been in any serious car accidents or have you ever been a victim of a crime? (Tell me about that.)

IF NO SUCH EVENTS, CIRCLE 1 AND GO TO ***PANIC DISORDER*** ON PAGE E.9.

IF ONE OR MORE SUCH EVENTS, CIRCLE 3 AND CONTINUE: **TRAUMATIC EVENT(S)** NO 1 YES 3 E1

Traumatic Events List

Brief Description	Date (Month/Yr)	Age
_____	____/____	___
_____	____/____	___
_____	____/____	___
_____	____/____	___
_____	____/____	___
_____	____/____	___
_____	____/____	___

IF ANY EVENTS LISTED: Sometimes traumatic experiences like (TRAUMAS LISTED ABOVE) keep coming back in nightmares, flashbacks, or thoughts that you can't get rid of. Has that ever happened to you?

IF NO: What about being very upset when you were in a situation that reminded you of one of these terrible things?

IF NO TO BOTH OF THE ABOVE, CIRCLE 1 AND GO TO ***PANIC DISORDER*** ON PAGE E.9. 1 3 E2

IF YES TO EITHER OR BOTH OF THE ABOVE, CIRCLE 3 AND CONTINUE:

POSTTRAUMATIC STRESS DISORDER CRITERIA

FOR FOLLOWING QUESTIONS, FOCUS ON TRAUMATIC EVENT(S) MENTIONED IN SCREENING QUESTION ABOVE.

A. The person has been exposed to a traumatic event in which both of the following were present:

IF MORE THAN ONE TRAUMA IS REPORTED: Which of these do you think affected you the most?

_____

_____

_____

(1) the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others — ? 1 2 3 — E3

GO TO ***Panic***, E.9

IF UNCLEAR: How did you react when (TRAUMA) happened? (Were you very afraid or did you feel helpless or horrified?)

_____

_____

(2) the person's response involved intense fear, helplessness or horror. — ? 1 2 3 — E4

GO TO ***Panic***, E.9

Now I'd like to ask a few questions about specific ways that it may have affected you in the past year.

For example, in the past year . . .

B. The traumatic event is persistently re-experienced in one (or more) of the following ways:

. . . did you think about (TRAUMA) when you didn't want to or did thoughts about (TRAUMA) come to you suddenly when you didn't want them to?

_____

_____

(1) recurrent and intrusive distressing recollections of the event, including images, thoughts or perceptions — ? 1 2 3 — E5

. . . what about having dreams about (TRAUMA)?

_____

_____

(2) recurrent distressing dreams of the event — ? 1 2 3 — E6

. . .what about finding yourself acting or feeling as if you were back in the situation? _____ _____ _____	(3)	acting or feeling as if the traumatic event were recurring (includes a sense of reliving the experience, illusions, hallucinations and dissociative flashback episodes, including those that occur on awakening or when intoxicated)	?	1	2	3	E7
. . . what about getting very upset when something reminded you of (TRAUMA)? _____ _____ _____	(4)	intense psychological distress at exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event	?	1	2	3	E8
. . . what about having physical symptoms--like breaking out in a sweat, breathing heavily or irregularly, or your heart pounding or racing, when something reminded you of (TRAUMA)? _____ _____ _____	(5)	physiological reactivity on exposure to internal or external cues that symbolize or resemble an aspect of the traumatic event	?	1	2	3	E9
	AT LEAST ONE “B” SX IS CODED “3”			1		3	E10

GO TO ***Panic***, E.9

C. Persistent avoidance of stimuli associated with the trauma and numbing of general responsiveness (not present before the trauma), as indicated by three (or more) of the following:

IF TRAUMA HAS OCCURRED IN THE PAST YEAR: Since (THE TRAUMA)...

IF TRAUMA OCCURRED PRIOR TO PAST YEAR: In the past year, that is since (CURRENT DATE) 2011

. . have you made a special effort to avoid thinking or talking about what happened?	(1) efforts to avoid thoughts, feelings, or conversations associated with the trauma	?	1	2	3	E11
. . . have you stayed away from things or people that reminded you of (TRAUMA)?	(2) efforts to avoid activities, places, or people that arouse recollections of the trauma	?	1	2	3	E12
. . . have you been unable to remember some important part of what happened?	(3) inability to recall an important aspect of the trauma	?	1	2	3	E13
. . . have you been less interested in doing things that used to be important to you, like seeing friends, reading books or watching TV?	(4) markedly diminished interest or participation in significant activities	?	1	2	3	E14

. . . have you felt distant or cut off from others?	(5)	feeling of detachment or estrangement from others	?	1	2	3	E15
. . . have you felt "numb" or like you no longer had strong feelings about anything or loving feelings for anyone?	(6)	restricted range of affect, (e.g., unable to have loving feelings)	?	1	2	3	E16
. . . did you notice a change in the way you think about or plan for the future?	(7)	sense of a foreshortened future (e.g., does not expect to have a career, marriage, children, or a normal life span)	?	1	2	3	E17
	AT LEAST 3 "C" SXS ARE CODED "3"			1 GO TO ***Panic***, E.9		3	E18

IF TRAUMA HAS OCCURRED IN THE PAST YEAR: Since (THE TRAUMA). . .

IF TRAUMA OCCURRED PRIOR TO PAST YEAR: In the past year…

D. Persistent symptoms of increased arousal (not present before the trauma) as indicated by two (or more) of the following:

. . . have you had trouble sleeping? (What kind of trouble?)	(1) difficulty falling or staying asleep	?	1	2	3	E19
. . . have you been unusually irritable? What about outbursts of anger?	(2) irritability or outbursts of anger	?	1	2	3	E20
. . . have you had trouble concentrating?	(3) difficulty concentrating	?	1	2	3	E21
. . . have you been watchful or on guard even when there was no reason to be?	(4) hypervigilance	?	1	2	3	E22
. . . have you been jumpy or easily startled, like by sudden noises?	(5) exaggerated startle response	?	1	2	3	E23
	AT LEAST TWO “D” SXS ARE CODED “3”		1		3	E24

GO TO ***Panic***, E.9

About how long did these problems--(CITE POSITIVE PTSD SYMPTOMS)--last?

_____

_____

_____

_____

E. Duration of the disturbance (symptoms in criteria B, C, and D) is more than one month — ? 1 2 3 — E25

1 → GO TO ***Panic***, E. 9

F. The disturbance causes clinically significant distress or impairment in social, occupational, or other important areas of functioning — ? 1 2 3 — E26

1 → GO TO ***Panic***, E.9

POSTTRAUMATIC STRESS DISORDER CRITERIA A, B, C, D, E, AND F ARE CODED "3" AND PRESENT IN THE PAST YEAR — 1 3 — E27

1 → GO TO ***PANIC***, E.9

3 → POST-TRAUMATIC STRESS DISORDER IN THE PAST YEAR

This page has been intentionally left blank.

PANIC DISORDER

PANIC DISORDER CRITERIA

IF SCREENING QUESTION #1 EQUALS 1, CIRCLE 1 AND GO TO ***AWOPD*,** ON PAGE E.15 E28

IF SCREENING QUESTION #1 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:
You've said that in the past year you have had a panic attack, when you suddenly felt frightened, or anxious or suddenly developed a lot of physical symptoms . . .

SCREEN Q#1	
NO	YES
1	3

Have these attacks ever come on completely out of the blue--in situations where you didn't expect to be nervous or uncomfortable?

IF UNCLEAR: How many of these kinds of attacks have you had? (At least two?)

A. (1) recurrent unexpected panic attacks. ? 1 2 3 E29

GO TO ***AWOPD***, E.15

After any of these attacks . . .

(2) at least one of the attacks has been followed by a month (or more) of one of the following: ? 1 2 3 E30

GO TO ***AWOPD***, E.15

Did you worry that there might be something terribly wrong with you, like you were having a heart attack or were going crazy? (How long did you worry?) (At least a month?)

(b) worry about the implications of the attack or its consequences (e.g., losing control, having a heart attack, "going crazy");

IF NO: Did you worry a lot about having another one? (How long did you worry?) (At least a month?)

(a) persistent concern about having additional attacks;

IF NO: Did you do anything differently because of the attacks (like avoiding certain places or not going out alone?) (What about avoiding certain activities like exercise?) (What about things like always making sure you're near a bathroom or exit?)

(c) a significant change in behavior related to the attacks;

NOW CHECK TO SEE IF CRITERIA ARE MET FOR A PANIC ATTACK.

When was the last bad one? What was the first thing you noticed? Then what?

_____

IF UNKNOWN: Did the symptoms come on all of a sudden? IF YES: How long did it take from when it began to when it got really bad? (Less than ten minutes?)		The panic attack symptoms developed abruptly and reached a peak within ten minutes	?	1 GO TO ***AWOPD***, E.15	2	3	E31
During that attack . . .							
. . .did your heart race, pound or skip?	(1)	palpitations, pounding heart, or accelerated heart rate	?	1	2	3	E32
. . .did you sweat?	(2)	sweating	?	1	2	3	E33
. . .did you tremble or shake?	(3)	trembling or shaking	?	1	2	3	E34
. . .were you short of breath? (Did you have trouble catching your breath?)	(4)	sensations of shortness of breath or smothering	?	1	2	3	E35
. . .did you feel as if you were choking?	(5)	feeling of choking	?	1	2	3	E36
. . .did you have chest pain or pressure?	(6)	chest pain or discomfort	?	1	2	3	E37
. . .did you have nausea or upset stomach or the feeling that you were going to have diarrhea?	(7)	nausea or abdominal distress	?	1	2	3	E38

. . .did you feel dizzy, unsteady, or like you might faint?	(8)	feeling dizzy, unsteady, light-headed or faint	?	1	2	3	E39
. . .did things around you seem unreal or did you feel detached from things around you or detached from part of your body?	(9)	derealization (feelings of un-reality) or depersonalization (being detached from oneself)	?	1	2	3	E40
. . .were you afraid you were going crazy or might lose control?	(10)	fear of losing control or going crazy	?	1	2	3	E41
. . .were you afraid that you might die?	(11)	fear of dying	?	1	2	3	E42
. . .did you have tingling or numbness in parts of your body?	(12)	paresthesias (numbness or tingling sensations)	?	1	2	3	E43
. . .did you have flushes (hot flashes) or chills?	(13)	chills or hot flushes	?	1	2	3	E44
		AT LEAST FOUR ITEMS CODED "3" AND REACHED A PEAK WITHIN 10 MINUTES (E31 CODED "3")		1 GO TO ***AWOPD*,** E.15		3	E45

Just before you began having panic attacks, were you taking any drugs, caffeine, diet pills, or other medicines?

(How much coffee, tea, or caffeinated soda do you drink a day?)

Just before the attacks, were you physically ill?

IF YES: What did the doctor say?

C. Not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition

? 1 3 E46

DUE TO SUBSTANCE USE OR GMC

GO TO ***AWOPD***, E.15

IF THERE IS ANY INDICATION THAT PANIC ATTACKS MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***ANXIETY DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET AND RETURN HERE TO MAKE A RATING OF "1" OR "3"

PRIMARY ANXIETY DISORDER

Etiological general medical conditions include: hyperthyroidism, hyperparathyroidism, pheochromocytoma, vestibular dysfunctions, seizure disorders, and cardiac conditions (e.g., arrhythmias, supraventricular tachycardia).

Etiological substances include: intoxication with central nervous stimulants (e.g., cocaine, amphetamines, caffeine) or cannabis or withdrawal from central nervous system depressants (e.g., alcohol, barbiturates) or from cocaine.

CONTINUE

D. The panic attacks are not better accounted for by another mental disorder, such as Social Phobia (e.g., occurring on exposure to feared social situations), Specific Phobia, Obsessive-Compulsive Disorder (e.g., on exposure to dirt in someone with an obsession about contamination), Posttraumatic Stress Disorder, or Separation Anxiety Disorder.

? 1 3 E47

A, C, AND D coded "3"

1 3 E47a

GO TO ***AWOPD***, E.15

PANIC DISORDER

PANIC DISORDER WITH AGORAPHOBIA

IF NOT OBVIOUS FROM OVERVIEW: Are there situations that make you nervous because you are afraid that you might have a panic attack?

Tell me about that.

IF CANNOT GIVE SPECIFICS: What about . . .

. . .being uncomfortable if you're more than a certain distance from home?
. . .being in a crowded place like a busy store, movie theatre, or restaurant?
. . .standing in a line?
. . .being on a bridge?
. . .using public transportation--like a bus, train, or subway--or driving a car?

_____

_____

_____

B. The presence of Agoraphobia: ? 1 2 3 E48

(1) Anxiety about being in places or situations from which escape might be difficult (or embarrassing) or in which help may not be available in the event of having an unexpected or situationally predisposed Panic Attack or panic-like symptoms. Agoraphobic fears typically involve characteristic clusters of situations that include being outside the home alone; being in a crowd or standing in a line; being on a bridge; and traveling in a bus, train or automobile.

PANIC DISORDER WITHOUT AGORA-PHOBIA

GO TO ***SOCIAL PHOBIA***, E.19

Do you avoid these situations?

IF NO: When you are in one of these situations, do you feel very uncomfortable or like you might have a panic attack?

(Can you go into one of these situations only if you are with someone you know?)

_____

_____

_____

_____

(2) Agoraphobic situations are avoided (e.g., travel is restricted), or else endured with marked distress or with anxiety about having a panic attack or panic-like symptoms, or require the presence of a companion. ? 1 2 3 E49

PANIC DISORDER WITHOUT AGORA-PHOBIA

GO TO ***SOCIAL PHOBIA***, E.19

(3) The anxiety or phobic avoidance is not better accounted for by another mental disorder, such as Social Phobia (e.g., avoidance limited to social situations because of fear of embarrassment), Specific Phobia (e.g., avoidance limited to a single situation like elevators), Obsessive-Compulsive Disorder (e.g., avoidance of dirt in someone with an obsession about contamination), Posttraumatic Stress Disorder (e.g., avoidance of stimuli associated with a severe stressor), or Separation Anxiety Disorder (e.g., avoidance of leaving home or relatives). ? 1 3 E50

1 → PANIC DISORDER WITHOUT AGORA-PHOBIA

GO TO ***SOCIAL PHOBIA***, E.19

NOTE: CONSIDER SPECIFIC PHOBIA IF FEAR IS LIMITED TO ONE OR ONLY A FEW SPECIFIC SITUATIONS OR SOCIAL PHOBIA IF FEAR IS LIMITED TO SOCIAL SITUATIONS

B(1), B(2), B(3) ALL CODED "3" 1 3 E51

1 → PANIC DISORDER WITHOUT AGORA-PHOBIA IN PAST YEAR

3 → PANIC DISORDER WITH AGORA-PHOBIA IN PAST YEAR

AGORAPHOBIA WITHOUT HISTORY OF PANIC DISORDER (AWOPD)

AGORAPHOBIA WITHOUT HISTORY OF PANIC DISORDER (AWOPD) CRITERIA

IF MET PAST YEAR CRITERIA FOR PANIC DISORDER, CIRCLE 3 AND GO TO ***SOCIAL PHOBIA*** ON PAGE E.19.

IF CRITERIA FOR PAST YEAR PANIC DISORDER NOT MET, CIRCLE 1 AND CONTINUE:

PANIC DISORDER	
NO	YES
1	3

E52

IF SCREENING QUESTION #2 EQUALS 1, CIRCLE 1 AND GO TO ***SOCIAL PHOBIA*** ON PAGE E.19.

IF SCREENING QUESTION #2 EQUALS 2 or 3, CIRCLE 3 AND CONTINUE:

SCREEN Q#2	
NO	YES
1	3

E53

You've said that in the past year you have been afraid of going out of the house alone, being in crowds, standing in a line, or traveling on buses or trains . . .

What were you afraid could happen?

_____

_____

_____

_____

_____

A. The presence of Agoraphobia:

(1) anxiety about being in places or situations from which escape might be difficult (or embarrassing) or in which help may not be available <u>in the event of having panic-like symptoms</u> (e.g., dizziness or diarrhea). Agoraphobic fears typically involve characteristic clusters of situations that include being outside the home alone; being in a crowd or standing in a line; being on a bridge; and traveling in a bus, train, or car.

? 1 2 3

GO TO ***SOCIAL PHOBIA***, E.19

E54

INDICATE FEARED SYMPTOM:

Do you avoid these situations?

IF NO: When you are in one of these situations, do you feel very uncomfortable or like you might have a panic attack?

(Can you go into one of these situations only if you are with someone you know?)

_____

_____

(2) Agoraphobic situations are avoided (e.g., travel is restricted), or else endured with marked distress or with anxiety about having panic-like symptoms, or require the presence of a companion.

? 1 2 3 E55

GO TO ***SOCIAL PHOBIA***, E. 19

(3) The anxiety or phobic avoidance is not better accounted for by another mental disorder, such as Social Phobia (e.g., avoidance limited to social situations because of fear of embarrassment), Specific Phobia (e.g., avoidance limited to single situations like elevators), Obsessive-Compulsive Disorder (e.g., avoidance of dirt in someone with an obsession about contamination), Posttraumatic Stress Disorder (e.g., avoidance of stimuli associated with a severe stressor), Separation Anxiety Disorder (e.g., avoidance of leaving home or relatives).

? 1 3 E56

GO TO ***SOCIAL PHOBIA***, E.19

NOTE: CONSIDER SPECIFIC PHOBIA IF FEAR IS LIMITED TO ONE OR ONLY A FEW SPECIFIC SITUATIONS, OR SOCIAL PHOBIA IF FEAR IS LIMITED TO SOCIAL SITUATIONS

A(1), A(2), A(3) ALL CODED "3" 1 3 E57

GO TO ***SOCIAL PHOBIA***, E.19

Just before you began having these fears, were you taking any drugs, caffeine, diet pills, or other medicines?

(How much coffee, tea, or caffeinated soda do you drink a day?)

Just before the fears began, were you physically ill?

IF YES: What did the doctor say?

_____

_____

C. Not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition

? 1 3 E58

DUE TO SUBSTANCE USE OR GMC

GO TO ***SOCIAL PHOBIA***, E.19

PRIMARY ANXIETY DISORDER

IF THERE IS ANY INDICATION THAT THE ANXIETY MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***ANXIETY DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, ND RETURN HERE TO MAKE A RATING OF "1" OR "3."

<u>Etiological general medical conditions include</u> hyper- and hypo-thyroidism, hypoglycemia, hyper-parathyroidism, pheochromocytoma, congestive heart failure, arrhythmias, pulmonary embolism, chronic obstructive pulmonary disease, pneumonia, hyperventilation, B-12 deficiency, porphyria, CNS neoplasms, vestibular dysfunction, encephalitis.

<u>Etiological substances include</u> intoxication with central nervous stimulants (e.g., cocaine, amphetamines, caffeine) or cannabis, hallucinogens, PCP, or alcohol, or withdrawal from central nervous system depressants (e.g., alcohol, sedatives, hypnotics) or from cocaine.

CONTINUE

D. If an associated general medical condition is present, the fear described in criterion A is clearly in excess of that usually associated with the condition.

1 3 E59

GO TO ***SOCIAL PHOBIA***, E. 19

AWOPD IN PAST YEAR

This page has been intentionally left blank.

SOCIAL PHOBIA

SOCIAL PHOBIA CRITERIA

IF SCREENING QUESTION #3 EQUALS 1, CIRCLE 1 AND GO TO ***SPECIFIC PHOBIA***, ON PAGE E.23.

IF SCREENING QUESTION #3 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:

SCREEN Q#3	
NO	YES
1	3

E60

You've said that during the past year there have been things that you are afraid to do in front of other people, like speaking, eating, or writing . . .

Tell me about it.

What are you afraid would happen when _____?

IF PUBLIC SPEAKING ONLY: (Do you think that you are more uncomfortable than most people are in that situation?)

_____

_____

_____

_____

_____

A. A marked and persistent fear of one or more social or performance situations in which the person is exposed to unfamiliar people or to possible scrutiny by others. The individual fears that he or she will act in a way (or show anxiety symptoms) that will be humiliating or embarrassing.

? 1 2 3 E61

GO TO ***SPECIFIC PHOBIA***, E.23

Note: In adolescents, there must be evidence of capacity for age-appropriate relationships with familiar people and the anxiety must occur in peer settings, not just in interactions with adults.

Have you always felt anxious when you (CONFRONTED PHOBIC STIMULUS)?

_____

_____

_____

B. Exposure to the feared social situation almost invariably provokes anxiety, which may take the form of a situationally bound or situationally predisposed panic attack.

? 1 2 3 E62

GO TO ***SPECIFIC PHOBIA***, E.23

Did you think that you are more afraid of (PHOBIC ACTIVITY) than you should have been (or than made sense)? _____	C. The person recognizes that the fear is excessive or unreasonable. Note: in children, this feature may be absent.	?	1	2	3	E63
			GO TO ***SPECIFIC PHOBIA***, E.23			
IF NOT OBVIOUS: Do you go out of your way to avoid _____? IF NO: How hard was it for you to _____? _____ _____	D. The feared social or performance situations are avoided, or else endured with intense anxiety or distress.	?	1	2	3	E64
			GO TO ***SPECIFIC PHOBIA***, E.23			
IF UNCLEAR WHETHER FEAR WAS CLINICALLY SIGNIFICANT: How much does _____ interfere with your life? IF DOES NOT INTERFERE WITH LIFE: How much has the fact that you have this fear bothered you? _____ _____	E. The avoidance, anxious anticipation, or distress in the feared social or performance situation(s) interferes significantly with the person's normal routine, occupational (academic) functioning, or social activities or relationships, or there is marked distress about having the phobia.	?	1	2	3	E65
			GO TO ***SPECIFIC PHOBIA***, E.23			
	F. NOTE: CRITERION F HAS BEEN OMITTED FROM THIS VERSION OF THE SCID.					

Just before you began having these fears, were you taking any drugs, caffeine, diet pills, or other medicines?

(How much coffee, tea, or caffeinated soda did you drink a day?)

Just before the fears began, were you physically ill?

IF YES: What did the doctor say?

G. The fear or avoidance is not due to the direct physiological effects of a substance (e.g., a drug of abuse, a medication) or a general medical condition.

? 1 3 E66

DUE TO SUBSTANCE USE OR GMC

GO TO ***SPECIFIC PHOBIA*,** E.23

PRIMARY ANXIETY DISORDER

IF THERE IS ANY INDICATION THAT THE ANXIETY MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF THE GMC OR SUBSTANCE), GO TO ***ANXIETY DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

<u>Etiological general medical conditions include:</u> hyper- and hypo-thyroidism, hypoglycemia, hyper-parathyroidism, pheochromocytoma, congestive heart failure, arrhythmias, pulmonary embolism, chronic obstructive pulmonary disease, pneumonia, hyperventilation, B-12 deficiency, porphyria, CNS neoplasms, vestibular dysfunction, encephalitis.

<u>Etiological substances include:</u> intoxication with central nervous stimulants (e.g., cocaine, amphetamines, caffeine) or cannabis, hallucinogens, PCP, or alcohol, or withdrawal from central nervous system depressants (e.g., alcohol, sedatives, hypnotics) or from cocaine.

CONTINUE

. . . and is not better accounted for by another mental disorder (e.g., Panic Disorder Without Agoraphobia, Separation Anxiety Disorder, Body Dysmorphic Disorder, a Pervasive Developmental Disorder, or Schizoid Personality Disorder).

? 1 2 3 E67

GO TO ***SPECIFIC PHOBIA***, E.23

IF NOT ALREADY CLEAR: RETURN TO THIS ITEM AFTER COMPLETING INTERVIEW.

H. If a general medical condition or other mental disorder is present, the fear in A is unrelated to it, e.g., the fear is not of stuttering, trembling (in Parkinson's disease) or exhibiting abnormal eating behavior (in Anorexia Nervosa or Bulimia Nervosa).

? 1 2 3 E68

1 → GO TO ***SPECIFIC PHOBIA***, E.23

SOCIAL PHOBIA CRITERIA A, B, C, D, E, G, AND H ARE CODED "3"

1 3 E69

1 → GO TO ***SPECIFIC PHOBIA***, E.23

3 → SOCIAL PHOBIA IN PAST YEAR

SPECIFIC PHOBIA **SPECIFIC PHOBIA CRITERIA**

IF SCREENING QUESTION #4 EQUALS 1, CICLE 1 AND GO TO ***OCD/OBSESSIONS*** ON PAGE E.27

IF SCREENING QUESTION #4 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:

SCREEN Q#4	
NO	YES
1	3

E70

You've said that in the past year there have been other things that you've been especially afraid of, like flying, seeing blood, getting a shot, heights, closed places, or certain kinds of animals or insects . . .

Tell me about that.

What are you afraid would happen when (CONFRONTED WITH PHOBIC STIMULUS)?

_____

_____

_____

_____

A. Marked and persistent fear that is excessive or unreasonable, cued by the presence or anticipation of a specific object or situation (e.g., flying, heights, animals, receiving an injection, seeing blood). ? 1 2 3 E71

GO TO ***OCD/ OBSESSION***, E.27

Have you always felt frightened when you (CONFRONTED PHOBIC STIMULUS)?

_____

_____

_____

B. Exposure to the phobic stimulus almost invariably provokes an immediate anxiety response, which may take the form of a situationally bound or situationally predisposed Panic Attack. ? 1 2 3 E72

GO TO *** OCD/ OBSESSION***, E.27

Did you think that you are more afraid of (PHOBIC STIMULUS) than you should have been (or than made sense)?

_____

_____

_____

_____

C. The person recognizes that the fear is excessive or unreasonable. ? 1 2 3 E73

GO TO ***OCD/ OBSESSION***, E.27

Do you go out of your way to avoid (PHOBIC STIMULUS)?

(Are there things you don't do because of this fear that you would otherwise have done?)

IF NO: How hard is it for you to (CONFRONT PHOBIC STIMULUS)?

_____

_____

_____

D. The phobic situation(s) is avoided, or else endured with intense anxiety or distress.

? 1 2 3 E74

GO TO ***OCD/ OBSESSION***, E.27

IF UNCLEAR WHETHER FEAR WAS CLINICALLY SIGNIFICANT: How much does (PHOBIA) interfere with your life?

(Is there anything you've avoided because of being afraid of [PHOBIC STIMULUS])?

IF DOES NOT INTERFERE WITH LIFE: How much has the fact that you were afraid of (PHOBIC STIMULUS) bothered you?

_____

_____

_____

_____

E. The avoidance, anxious anticipation, or distress in the feared situation(s) interferes significantly with the person's normal routine, occupational (or academic) functioning, or social activities or relationships, or there is marked distress about having the phobia.

? 1 2 3 E75

GO TO ***OCD/ OBSESSION***, E.27

NOTE: CRITERION F HAS BEEN OMITTED FROM THIS VERSION OF THE SCID.

IF NOT ALREADY CLEAR: RETURN TO THIS ITEM AFTER COMPLETING SECTION ON PTSD AND OBSESSIVE-COMPULSIVE DISORDER.

G. The anxiety, panic attacks, or phobic avoidance associated with the specific object or situation are not better accounted for by another mental disorder, such as Obsessive-Compulsive Disorder (e.g., fear of dirt in someone with an obsession about contamination), Posttraumatic Stress Disorder (e.g. avoidance of stimuli associated with a severe stressor), Separation Anxiety Disorder (e. g., avoidance of school), Social Phobia (e.g., avoidance of social situations because of fear of embarrassment), Panic Disorder With Agoraphobia, or Agoraphobia Without History of Panic Disorder.

? 1 3 E76

GO TO * **OCD/ OBSESSION***, E.27

SPECIFIC PHOBIA CRITERIA A, B, C, D, E, AND G ARE CODED "3."

1 3 E77

GO TO * **OCD/ OBSESSION***, E.27

SPECIFIC PHOBIA PAST YEAR

This page has been intentionally left blank.

OBSESSIVE COMPULSIVE DISORDER

OBSESSIVE COMPULSIVE DISORDER CRITERIA

SCREEN Q#5	
NO	YES
1	3

E78

IF SCREENING QUESTION #5 EQUALS 1, CIRCLE 1 AND GO TO ***COMPULSIONS*** ON PAGE E.29

IF SCREENING QUESTION #5 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE: You've said that in the past year that you have had thoughts that didn't make any sense and kept coming back to you even when you tried not to have them . . .

A. Either obsessions or compulsions:

(What were they?)

Obsessions as defined by (1), (2), (3) and (4)

IF SUBJECT NOT SURE WHAT IS MEANT: . . . Thoughts like hurting someone, even though you really didn't want to or being contaminated by germs or dirt?

_____

_____

_____

_____

(1) recurrent and persistent thoughts, impulses, or images that are experienced, at some time during the disturbance, as intrusive and inappropriate, and that cause marked anxiety or distress — ? 1 2 3 — E79

(2) the thoughts, impulses, or images are not simply excessive worries about real-life problems. — ? 1 2 3 — E80

When you had these thoughts, did you try hard to get them out of your head? (What would you try to do?)

_____

_____

(3) the person attempts to ignore or suppress such thoughts, impulses, or images, or to neutralize them with some other thought or action. — ? 1 2 3 — E81

IF UNCLEAR: Where did you think these thoughts were coming from?

_____

(4) the person recognizes that the obsessional thoughts, impulses, or images are a product of his or her own mind (not imposed from without as in thought insertion) — ? 1 2 3 — E82

OBSES-SIONS

NO OBSESSIONS GO TO ***COMPULSIONS***, E.29

DESCRIBE CONTENT OF OBSESSIONS:

_____

This page has been intentionally left blank.

COMPULSIONS

SCREEN Q#6	
NO	YES
1	3

E83

IF SCREENING QUESTION #6 EQUALS 1, CIRCLE 1 AND GO TO ***CHECK FOR OBSESSIONS/COMPULSIONS*** ON PAGE E.30 (TOP OF NEXT PAGE)

IF SCREENING QUESTION #6 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:
You've said that in that past year there were things that you had to do over and over again and couldn't resist doing, like washing your hands again and again, counting up to a certain number or checking something several times to make sure that you had done it right . . .

(What did you have to do?)

_____

_____

_____

_____

Compulsions as defined by (1) and (2)

(1) repetitive behaviors (e. g., handwashing, ordering, checking) or mental acts (e.g., praying, counting, repeating words silently) that the person feels driven to perform in response to an obsession, or according to rules that must be applied rigidly — ? 1 2 3 — E84

IF UNCLEAR: Why did you have to do (COMPULSIVE ACT?) What would happen if you didn't do it?

IF UNCLEAR: How many times would you do (COMPULSIVE ACT)? How much time a day would you spend doing it?

_____

_____

(2) the behaviors or mental acts are aimed at preventing or reducing distress or preventing some dreaded event or situation; however these behaviors or mental acts either are not connected in a realistic way with what they are designed to neutralize or prevent, or are clearly excessive — ? 1 2 3 — E85

COMPULSIONS

GO TO ***CHECK FOR OBSESSIONS / COMPULSIONS***, E.30 (TOP OF NEXT PAGE)

DESCRIBE CONTENT OF COMPULSION(S)

_____

_____

_____

CHECK FOR OBSESSIONS/COMPULSIONS

IF NEITHER **OBSESSIONS** NOR **COMPULSIONS**, CIRCLE 1 AND GO TO ***GENERALIZED ANXIETY*** ON PAGE E.33

IF EITHER OBSESSIONS, OR COMPULSIONS, OR BOTH, CIRCLE 3 AND CONTINUE:

OBSESSIONS/ COMPULSIONS	
NO	YES
1	3

E86

Do you (think about [OBSESSIVE THOUGHTS]/do [COMPULSIVE ACTS]) more than you should have (or than makes sense)?

IF NO: How about when you first started having this problem?

_____

_____

_____

B. At some point during the course of the disorder, the person has recognized that the obsessions or compulsions are excessive or unreasonable. Note: this does not apply to children. ? 1 2 3 E87

GO TO ***GAD***, E.33

E88

Check here ___ if **With Poor Insight:** i.e., for most of the time during the current episode, the person does not recognize that the obsessions and compulsions are excessive or unreasonable.

What effect has this (OBSESSION OR COMPULSION) had on your life? (Did [OBSESSION OR COMPULSION] bother you a lot?)

_____

_____

(How much time have you spent on [OBSESSION OR COMPULSION])?

_____

_____

C. The obsessions or compulsions cause marked distress, are time-consuming (take more than an hour a day), or significantly interfere with the person's normal routine, occupational functioning, or usual social activities or relationships. ? 1 2 3 E89

GO TO ***GAD***, E.33

IF NOT ALREADY CLEAR: RETURN TO THIS ITEM AFTER COMPLETING INTERVIEW

D. If another Axis I disorder is present, the content of the obsessions or compulsions is not restricted to it (e.g., preoccupation with food in the presence of an Eating Disorder; hair pulling in the presence of Trichotillomania; concern with appearance in the presence of Body Dysmorphic Disorder; preoccupation with drugs in the presence of a Substance Use Disorder; preoccupation with having a serious illness in the presence of Hypochondriasis; preoccupation with sexual urges or fantasies in the presence of a Paraphilia, or guilty ruminations in the presence of Major Depressive Disorder).

? 1 3 E90

1 → GO TO ***GAD***, E.33

Just before you began having (OBSESSIONS OR COMPULSIONS) were you taking any drugs or medicines?

Just before the (OBSESSIONS OR COMPULSIONS) started, were you physically ill? (What did the doctor say?)

E. Not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition

? 1 3 E91

1 → DUE TO SUBSTANCE USE OR A GMC

GO TO ***GAD***, E.33

IF THERE IS ANY INDICATION THAT THE OBSESSIONS OR COMPULSIONS MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSEQUENCE OF A GMC OR SUBSTANCE), GO TO ***ANXIETY DUE TO GMC/SUBSTANCE,*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

3 → PRIMARY ANXIETY DISORDER

<u>Etiological general medical conditions include</u>: certain CNS neoplasms.

<u>Etiological substances include</u>: intoxication with central nervous system stimulants (e.g., cocaine, amphetamines)

CONTINUE

OBSESSIVE COMPULSIVE DISORDER CRITERIA A, B, C, D, AND E ARE CODED "3"

1 3 E92

1 → GO TO ***GAD***, E. 33

3 → OBSESSIVE COMPULSIVE DISORDER IN PAST YEAR

This page has been intentionally left blank.

GENERALIZED ANXIETY DISORDER

GENERALIZED ANXIETY DISORDER CRITERIA

IF SCREENING QUESTION #7 EQUALS 1, CIRCLE 1 AND GO TO *ANOREXIA* ON PAGE F.1.

IF SCRENING QUESTION #7 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:
You've said that in the last year there have been times you've been particularly nervous or anxious . . .

SCREEN Q#7	
NO	YES
1	3

E93

Do you also worry a lot about bad things that might happen? IF YES: What do you worry about? (How much do you worry about [EVENTS OR ACTIVITIES]?) Has there been a six month period of time in the past year when you were worrying for more days than not? _____ _____ _____	A. Excessive anxiety and worry (apprehensive expectation), occurring more days than not for at least six months, about a number of events or activities (such as work or school performance)	?	1	2	3	E94
When you're worrying this way, do you find that it's hard to stop yourself? _____ _____	B. The person finds it difficult to control the worry.	?	1	2	3	E95
When did this anxiety start? COMPARE ANSWER WITH ONSET OF MOOD OR PSYCHOTIC DISORDER. _____ _____ _____	F(2). Does not occur exclusively during the course of a Mood Disorder, Psychotic Disorder, or a Pervasive Developmental Disorder	?	1		3	E96

GO TO ***ANOREXIA***, F.1

Now I am going to ask you some questions about symptoms that often go along with being nervous.

Thinking about those periods in the past year when you're feeling nervous or anxious . . .

C. The anxiety and worry are associated with three (or more) of the following six symptoms (with at least some symptoms present for more days than not for the past six months):

. . . do you often feel physically restless--can't sit still?	(1)	restlessness or feeling keyed up or on edge	?	1	2	3	E97
. . . do you often feel keyed up or on edge?							
. . . do you often tire easily?	(2)	being easily fatigued	?	1	2	3	E98
. . . do you have trouble concentrating or does your mind go blank?	(3)	difficulty concentrating or mind going blank	?	1	2	3	E99
. . . are you often irritable?	(4)	irritability	?	1	2	3	E100
. . . are your muscles often tense?	(5)	muscle tension	?	1	2	3	E101

. . . do you often have trouble falling or staying asleep?	(6) sleep disturbance (difficulty falling or staying asleep, or restless unsatisfying sleep)	?	1	2	3	E102
	AT LEAST THREE "C" SXS ARE CODED "3"	?	1	2	3	E103
			GO TO ***ANOREXIA***, F.1			
CODE BASED ON PREVIOUS INFORMATION	D. The focus of the anxiety and worry is not confined to the features of another Axis I Disorder, e.g., the anxiety or worry is not about having a panic attack (as in Panic Disorder), being embarrassed in public (as in Social Phobia), being contaminated (as in Obsessive Compulsive Disorder), being away from home or close relatives (as in Separation Anxiety Disorder), gaining weight (as in Anorexia Nervosa), having multiple physical complaints (as in Somatization Disorder), or having a serious illness (as in Hypochondriasis), and the anxiety or worry do not occur exclusively during Posttraumatic Stress Disorder.	?	1		3	E104
			GO TO ***ANOREXIA***, F.1			
IF UNCLEAR: What effect has the anxiety, worry, or (PHYSICAL SYMPTOMS) had on your life? (Has it made it hard for you to do your work or be with your friends?)	E. The anxiety, worry, or physical symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning	?	1	2	3	E105
			GO TO ***ANOREXIA***, F.1			

Just before you began having this anxiety, were you taking any drugs, caffeine, diet pills, or other medicines?

(How much coffee, tea, or caffeinated soda do you drink a day?)

Just before these problems began, were you physically ill?

IF YES: What did the doctor say?

F. Not due to the direct physiological effects of a substance (e.g., a drug of abuse, medication) or to a general medical condition

? 1 3 E106

DUE TO SUB-STANCE USE OR A GMC

GO TO ***ANOREXIA***, F.1

IF THERE IS ANY INDICATION THAT THE ANXIETY MAY BE SECONDARY (I.E., A DIRECT PHYSIOLOGICAL CONSE-QUENCE OF A GMC OR SUB-STANCE), GO TO ***ANXIETY DUE TO GMC/SUBSTANCE*** IN THE BACK OF THIS BOOKLET, AND RETURN HERE TO MAKE A RATING OF "1" OR "3."

PRIMARY ANXIETY DISORDER

Etiological general medical conditions include: hyper- and hypo-thyroidism, hypoglycemia, hyper-parathyroidism, pheochromocytoma, congestive heart failure, arrhythmias, pulmonary embolism, chronic obstructive pulmonary disease, pneumonia, hyperventilation, B-12 deficiency, porphyria, CNS neoplasms, vestibular dysfunction, encephalitis.

Etiological substances include: intoxication with central nervous stimulants (e.g., cocaine, ampheta-mines, caffeine) or cannabis, hallu-cinogens, PCP, or alcohol or withdrawal from central nervous system depressants (e.g., alcohol, sedatives, hypnotics) or from cocaine.

CONTINUE

GENERALIZED ANXIETY CRITERIA A, B, C, D, E AND F ARE CODED "3"

1 3 E107

GO TO ***ANOREXIA***, F.1

GENERAL-IZED ANXIETY DISORDER

F. EATING DISORDERS

ANOREXIA NERVOSA **ANOREXIA NERVOSA CRITERIA**

IF SCREENING QUESTION #8 EQUALS 1, CIRCLE 1 AND GO TO ***BULIMIA*** ON PAGE F.3

IF SCREENING QUESTION #8 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:
You've said that there was a time in the past year
When you weighed much less than other people
thought you ought to weigh . . .

SCREEN Q#8	
NO	YES
1	3

F1

Why was that? How much did you weigh? How tall are you?

_____

_____

_____

A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected)

? 1 2 3 F2

GO TO ***BULIMIA***, F.3

At that time, were you very afraid that you could become fat?

_____

_____

B. Intense fear of gaining weight or becoming fat, even though underweight.

? 1 2 3 F3

GO TO ***BULIMIA***, F.3

At your lowest weight, did you still feel too fat or that part of your body was too fat?

IF NO: Did you need to be very thin in order to feel good about yourself?

IF NO AND LOW WEIGHT IS MEDICALLY SERIOUS: When you were that thin, did anybody tell you it could be dangerous to your health to be that thin? (What did you think?)

_____

_____

_____

C. Disturbance in the way in which one's body weight or shape is experienced; undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight

? 1 2 3 F4

GO TO ***BULIMIA***, F.3

FOR FEMALES: Before this time, were you having your periods? Did they stop? (For how long?)

_____

_____

_____

D. In postmenarchal females, amenorrhea, i.e., the absence of at least three consecutive menstrual cycles. (A woman is still considered to have amenorrhea if her periods occur only following hormone, e.g., estrogen, administration) ? 1 2 3 F5

GO TO ***BULIMIA***, F.3

ANOREXIA NERVOSA CRITERIA A, B, C, AND D ARE CODED "3" 1 3 F6

GO TO ***BULIMIA***, F.3

ANOREXIA NERVOSA PAST YEAR

BULIMIA NERVOSA **BULIMIA NERVOSA CRITERIA**

→ IF CRITERIA MET FOR ANOREXIA NERVOSA, CIRCLE 3 AND GO TO ***IED*** ON PAGE G.1.

→ IF CRITERIA NOT MET FOR ANOREXIA NERVOSA, CIRCLE 1 AND CONTINUE.

ANOREXIA NERVOSA	
NO	YES
1	3

F7

→ IF SCREENING QUESTION #9 EQUALS 1, CIRCLE 1 AND GO TO ***IED*** ON PAGE G.1.

→ IF QUESTION #9 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE:
You've said that in the past year, you've often had times when your eating was out of control. Tell me about those times.

SCREEN Q#9	
NO	YES
1	3

F8

A. Recurrent episodes of binge eating. An episode of binge eating is characterized by BOTH of the following:

(2) a sense of lack of control over eating during the episode (e.g., a feeling that one cannot stop eating or control what or how much one is eating)

? 1 2 3 F9

GO TO ***IED***, G.1

(1) eating, in a discrete period of time (e.g., within any two hour period), an amount of food that is definitely larger than most people would eat during a similar period of time and under similar circumstances.

? 1 2 3 F10

GO TO ***IED***, G.1

IF UNCLEAR: During these times, do you often eat within any two hour period what most people would regard as an unusual amount of food? Tell me about that.

B. Recurrent inappropriate compensatory behavior in order to prevent weight gain, such as: self-induced vomiting; misuse of laxatives, diuretics, enemas, or other medications; fasting; or excessive exercise.

? 1 2 3 F11

GO TO ***IED***, G.1

Did you do anything to counteract the effects of eating that much? (Like making yourself vomit, taking laxatives, enemas or water pills, strict dieting or fasting, or exercising a lot?)

How often were you eating that much (AND COMPENSATORY BEHAVIOR)? (At least twice a week for at least three months?)

_____

_____

_____

_____

C. The binge eating and inappropriate compensatory behaviors both occur, on average, at least twice a week for three months.

? 1 2 3 F12

GO TO ***IED***, G.1

Were your body shape and weight among the most important things that affected how you felt about yourself?

_____

_____

D. Self-evaluation is unduly influenced by body shape and weight.

? 1 2 3 F13

GO TO ***IED***, G.1

E. The disturbance does not occur exclusively during episodes of Anorexia Nervosa

? 1 3 F14

GO TO ***IED***, G.1

BULIMIA NERVOSA CRITERIA A, B, C, D AND E ARE CODED "3"

1 3 F15

GO TO ***IED***, G.1

BULIMIA NERVOSA PAST YEAR

G. INTERMITTENT EXPLOSIVE DISORDER

INTERMITTENT EXPLOSIVE DISORDER CRITERIA

In the past year, that is since (CURRENT DATE) 2011, have you had times when you lost control of your anger, resulting in your hitting or seriously threatening someone or damaging things?

IF YES: What did you do? When did you do it? How often did it happen?

DESCRIBE ASSAULTIVE ACTS:

_____

_____

A. Several discrete episodes of failure to resist aggressive impulses that result in serious assaultive acts or destruction of property. ? 1 2 3 G1

GO TO ***ALCOHOL USE DISORDERS***, H.1

What happened that set you off? (Do you think your reaction was much stronger than it should have been given the circumstances?) (Has anyone told you that your reaction was way off-base given the situation?)

_____

_____

B. The degree of aggressiveness expressed during the episodes is grossly out of proportion to any precipitating psychosocial stressors. ? 1 2 3 G2

GO TO ***ALCOHOL USE DISORDERS***, H.1

Did this happen only when you were drinking or using drugs?

Did this happen only when you were sick with a medical illness?

IF HX OF MANIA OR PSYCHOSIS: Did this happen only when you were feeling excited or irritable or only when you were (PSYCHOTIC SXS)?

(Did you do [ASSAULTIVE ACTS] because you were hearing voices or because your thinking was confused?)

(Did you do [ASSAULTIVE ACTS] on purpose or was it really beyond your control?)

_____

C. The aggressive episodes are not better accounted for by Antisocial Personality Disorder, Borderline Personality Disorder, a psychotic disorder, a Manic Episode, Conduct Disorder, or Attention-Deficit/ Hyperactivity Disorder and are not due to the physiological effects of a substance or a general medical condition. ? 1 2 3 G3

GO TO ***ALCOHOL USE DISORDERS***, H.1

G4

ITEMS A, B, AND C ARE CODED "3"

1 → GO TO ***ALCOHOL USE DISORDERS***, H.1

3 → INTER-MITTENT EXPLOSIVE DISORDER PAST YEAR

H. SUBSTANCE USE DISORDERS

ALCOHOL USE DISORDERS (PAST YEAR)

Next I'd like to ask about your use of alcohol. What have your drinking habits been like in the past year?
(How much do you drink?) (Have there been any times in the past year when you had five or more drinks on one occasion?)

RECORD TYPICAL WEEKLY PATTERN OF USE:

_____

_____

_____

_____

_____

When in the past year were you drinking the most? (How long did that period last?)

RECORD DATE OF HEAVIEST USE AND DESCRIBE PATTERN

_____

_____

_____

_____

_____

During that time . . .

how often were you drinking?

what were you drinking? how much?

During that time . . .

did your drinking cause problems for you?

did anyone object to your drinking?

_____

_____

_____

_____

_____

_____

_____

_____

IF R HAS NOT DRUNK AT LEAST 6 DRINKS IN THE PAST YEAR, CIRCLE THE 1 AND SKIP TO ***NON-ALCOHOL SUBSTANCE USE DISORDERS***, H. 9 1 3 H1

IF R HAS DRUNK AT LEAST 6 DRINKS IN THE PAST YEAR, CIRCLE THE 3 AND CONTINUE TO NEXT PAGE.

ALCOHOL DEPENDENCE

I'd now like to ask you some more questions about (TIME IN PAST YEAR WHEN DRINKING THE MOST OR TIME WHEN DRINKING CAUSED MOST PROBLEMS).

During that time…

ALCOHOL DEPENDENCE CRITERIA

A maladaptive pattern of alcohol use, leading to clinically significant impairment or distress, as manifested by three (or more) of the following occurring at any time in the same twelve month period:

NOTE: CRITERIA FOR ALCOHOL DEPENDENCE ARE NOT IN DSM-IV-TR ORDER

…did you often find that when you started drinking you ended up drinking much more than you were planning to? (Tell me about that.)

IF NO: What about drinking for a much longer period of time than you were planning to?

_____

_____

_____

(3) alcohol is often taken in larger amounts OR over a longer period than was intended ? 1 2 3 H2

…did you try to cut down or stop drinking alcohol?

IF YES: Did you actually stop drinking altogether?

(How many times did you try to cut down or stop altogether?)

IF NO: Did you want to stop or cut down? (Is this something you kept worrying about?)

_____

_____

_____

(4) there is a persistent desire OR unsuccessful efforts to cut down or control alcohol use ? 1 2 3 H3

…did you spend a lot of time drinking, being high, or hung over? (How much time?) _____ _____ _____	(5)	a great deal of time is spent in activities necessary to obtain alcohol, use alcohol, or recover from its effects	?	1	2	3	H4
…did you have times when you would drink so often that you started to drink instead of working or spending time at hobbies or with your family or friends, or engaging in other important activities, such as sports, gardening, or playing music? _____ _____ _____	(6)	important social, occupational, or recreational activities given up or reduced because of alcohol use	?	1	2	3	H5
IF NOT ALREADY KNOWN: During that time did your drinking cause any psychological problems like making you depressed or anxious, making it difficult to sleep, or causing "blackouts?" IF NOT ALREADY KNOWN: Did your drinking cause significant physical problems or made a physical problem worse? IF YES TO EITHER OF ABOVE: Did you keep on drinking anyway? _____ _____ _____	(7)	alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by alcohol (e.g., continued drinking despite recognition that an ulcer was made worse by alcohol consumption)	?	1	2	3	H6

Have you found that you needed to drink a lot more in order to get the feeling you wanted than you did when you first started drinking? IF YES: How much more? IF NO: What about finding that when you drank the same amount, it had much less effect than before? _____ _____ _____	(1) tolerance, as defined by either of the following: (a) a need for markedly increased amounts of alcohol to achieve intoxication or desired effect (b) markedly diminished effect with continued use of the same amount of alcohol	?	1	2	3	H7
During the past year have you had any withdrawal symptoms when you cut down or stopped drinking like . . . . . . sweating or racing heart? . . . hand shakes? . . . trouble sleeping? . . . feeling nauseated or vomiting? . . . feeling agitated? . . . or feeling anxious? (How about having a seizure or seeing, feeling, or hearing things that weren't really there?) _____ _____ IF NO: During the past year, have you ever started the day with a drink, or did you often drink or take some other drug or medication to keep yourself from getting the shakes or becoming sick? _____ _____	(2) withdrawal, as manifested by either (a) or (b): (a) at least <u>TWO</u> of the following: - - autonomic hyperactivity (e.g., sweating or pulse rate greater than 100) - - increased hand tremor - - insomnia - - nausea or vomiting - - psychomotor agitation - - anxiety - - grand mal seizures - - transient visual, tactile, or auditory hallucinations or illusions (b) alcohol (or a substance from the sedative / hypnotic / anxiolytic class) taken to relieve or avoid withdrawal symptoms	?	1	2	3	H8

AT LEAST THREE DEPENDENCE ITEMS CODED "3" AND ITEMS OCCURRED IN THE PAST YEAR

1

3

H9

ALCOHOL DEPENDENCE

GO TO ***ALCOHOL ABUSE***, H.6

GO TO ***ALCOHOL ABUSE***, H.6

PAST YEAR ALCOHOL ABUSE

Let me ask you a few more questions about (TIME IN PAST YEAR WHEN DRINKING THE MOST OR TIME WHEN DRINKING CAUSED MOST PROBLEMS).

During that time…

ALCOHOL ABUSE CRITERIA

A. A maladaptive pattern of alcohol use, leading to clinically significant impairment or distress, as manifested by one (or more) of the following occurring within a twelve month period:

…did you miss work or school because you were intoxicated, high, or very hung over? (What about doing a bad job at work or failing courses at school because of your drinking?) IF NO: What about not keeping your house clean [IF CHILDREN: or not taking proper care of your children] because of your drinking? IF YES TO EITHER: How often? (Over what period of time?) _____ _____ _____ _____	(1)	Recurrent alcohol use resulting in a failure to fulfill major role obligations at work, school, or home (e.g., repeated absences or poor work performance related to alcohol use; alcohol-related absences, suspensions, or expulsions from school; neglect of children or household).	?	1	2	3	H10
…did you drink in a situation in which it might have been dangerous to drink at all? (In the past year have you driven while you were really too drunk to drive?) IF YES AND UNKNOWN: How many times? (When?) _____ _____ _____	(2)	Recurrent alcohol use in situations in which it is physically hazardous (e.g., driving an automobile or operating a machine when impaired by alcohol use)	?	1	2	3	H11

During the past year has your drinking gotten you into trouble with the law? IF YES AND UNKNOWN: How often? (Over what period of time?) _____ _____ _____	(3)	recurrent alcohol-related legal problems (e.g., arrests for alcohol-related disorderly conduct)	?	1	2	3	H12
IF NOT ALREADY KNOWN: Did your drinking cause problems with other people, such as with family members, friends, or people at work? Did you get into physical fights when you were drinking? What about having bad arguments about what happens when you drink too much?) IF YES: Did you keep on drinking anyway? (Over what period of time?) _____ _____ _____	(4)	continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance (e.g., arguments with spouse about consequences of intoxication, physical fights)	?	1	2	3	H13
		AT LEAST ONE "A" ITEM CODED "3"		1		3	H14

GO TO ***NON-ALCOHOL SUBSTANCE USE DISORDERS***, H.9

ALCOHOL ABUSE

GO TO ***NON-ALCOHOL SUBSTANCE USE DISORDERS***, H.9

This page has been intentionally left blank.

NON-ALCOHOL SUBSTANCE USE DISORDERS
(PAST YEAR DEPENDENCE AND ABUSE)

Now I am going to ask you about your use of drugs or medicines in the past 12 months.

CIRCLE THE NAME OF EACH DRUG USED IN THE PAST YEAR (OR WRITE IN NAME IF "OTHER")	RECORD PATTERN OF USUAL USE AND PERIOD/PATTERN OF HEAVIEST USE (INCLUDING DATE AND DURATION)	INDICATE PAST YEAR USE LEVEL*	
In the past 12 months have you taken any pills to calm you down or mellow you out or to help you sleep - drugs like Valium, Xanax, Ativan, Klonopin, Rohypnol or “roofies”, Ambien, Sonata, Lunesta, Halcion, or Restoril?	**Sedatives-hypnotics-anxiolytics:** _____ _____ _____	1 3	H15
How about stimulants or “uppers”, like speed, methamphetamine, crystal meth, “crank”, Ritalin, dexadrine, Adderall or prescription diet pills?	**Stimulants:** _____ _____ _____	1 3	H16
How about prescription pain relievers like morphine, codeine, Darvocet, Darvon, Tylenol with Codeine, Percocet, Percodan, Tylox, Vicodin, Lortab, Lorcet, OxyContin, or any other prescription pain reliever?	**Opioids:** _____ _____ _____	1 3	H17
How about marijuana (pot, grass, weed) or hashish?	**Cannabis:** _____ _____	1 3	H18
How about heroin?	**Heroin:** _____	1 3	H19
How about cocaine, “crack”, or freebase?	**Cocaine:** _____	1 3	H20
How about LSD, “acid”, PCP, peyote, mescaline, psilocybin, Ecstasy, Ketamine or other hallucinogens?	**Hallucinogens/PCP** _____	1 3	H21
How about sniffing glue, paint, correction fluid, “poppers,” gasoline, laughing gas or other inhalants to get high?	**Inhalants:** _____	1 3	H22

*FOR ANY DRUG CLASS USED NONMEDICALLY MORE THAN ONCE (FOR CANNABIS, THRESHOLD IS AT LEAST 6 TIMES) IN THE PAST YEAR, CIRCLE 3 FOR USE LEVEL. FOR ALL OTHERS, CIRCLE 1.

*FOR PRESCRIBED MEDICATIONS, CIRCLE 3 IF SUBJECT REPORTS BEING DEPENDENT ON A PRESCRIBED DRUG OR USING MORE THAN WAS PRESCRIBED.

IF NO DRUG CLASSES HAVE A 3 CIRCLED FOR PAST YEAR USE LEVEL, CIRCLE 1 AND GO TO ***ADJUSTMENT DISORDER***, J.1. 1 3 H23

IF ANY DRUG CLASS HAS A 3 CIRCLED FOR LEVEL OF USE, CIRCLE 3 AND CONTINUE.

SUBSTANCE DEPENDENCE

I'd now like to ask you some more questions about (TIME IN THE PAST YEAR WHEN YOU WERE USING THE MOST DRUG[S] / YOUR USE OF DRUG[S] DURING THE PAST 12 MONTHS).

During that time...

SUBSTANCE DEPENDENCE CRITERIA

A maladaptive pattern of substance use, leading to clinically significant impairment or distress, as manifested by three (or more) of the following occurring at any time in the same twelve month period:

NOTE: CRITERIA FOR SUBSTANCE DEPENDENCE ARE NOT IN DSM-IV-TR ORDER

…did you often find that when you started using (DRUG[S]) you ended up using much more of it than you were planning to? (Tell me about it.) IF NO: What about using it over a much longer period of time than you were planning to? _____ _____ _____	(3) substance is often taken in larger amounts OR over a longer period than was intended	SED	?	1	2	3	H24
		STIM	?	1	2	3	H25
		OPI	?	1	2	3	H26
		CAN	?	1	2	3	H27
		HER	?	1	2	3	H28
		COC	?	1	2	3	H29
		HAL	?	1	2	3	H30
		INH	?	1	2	3	H31
…did you try to cut down or stop using (DRUG[S])? → IF YES: In the past year, did you ever actually stop using (DRUG[S]) altogether? (How many times did you try to cut down or stop altogether?) → IF NO: Did you want to stop or cut down? (Is this something you kept worrying about?) _____ _____ _____	(4) there is a persistent desire OR unsuccessful efforts to cut down or control substance use	SED	?	1	2	3	H32
		STIM	?	1	2	3	H33
		OPI	?	1	2	3	H34
		CAN	?	1	2	3	H35
		HER	?	1	2	3	H36
		COC	?	1	2	3	H37
		HAL	?	1	2	3	H38
		INH	?	1	2	3	H39

…did you spend a lot of time using (DRUG[S]) or doing whatever you had to do to get it? Did it take you a long time to get back to normal? (How much time?)	(5)	a great deal of time is spent in activities necessary to obtain the substance, use the substance, or recover from its effects	SED	?	1	2	3	H40
			STIM	?	1	2	3	H41
			OPI	?	1	2	3	H42
			CAN	?	1	2	3	H43
			HER	?	1	2	3	H44
			COC	?	1	2	3	H45
			HAL	?	1	2	3	H46
			INH	?	1	2	3	H47
…did you often have times when you would use (DRUG[S]) so often that you used (DRUG[S]) instead of working or spending time with your family or friends or engaging in other important activities, such as sports, gardening, or playing music?	(6)	important social, occupational, or recreational activities given up or reduced because of substance use	SED	?	1	2	3	H48
			STIM	?	1	2	3	H49
			OPI	?	1	2	3	H50
			CAN	?	1	2	3	H51
			HER	?	1	2	3	H52
			COC	?	1	2	3	H53
			HAL	?	1	2	3	H54
			INH	?	1	2	3	H55
IF NOT ALREADY KNOWN: Did [DRUG(S)] cause any psychological problems like making you depressed, agitated, or paranoid? IF NOT ALREADY KNOWN: Did [DRUG(S)] cause any significant physical problems or make a physical problem worse? IF YES TO EITHER OF ABOVE: Did you keep on using (DRUG[S]) anyway?	(7)	substance use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by the substance (e.g., recurrent cocaine use despite recognition of cocaine-related depression)	SED	?	1	2	3	H56
			STIM	?	1	2	3	H57
			OPI	?	1	2	3	H58
			CAN	?	1	2	3	H59
			HER	?	1	2	3	H60
			COC	?	1	2	3	H61
			HAL	?	1	2	3	H62
			INH	?	1	2	3	H63

Have you found that you needed to use a lot more (DRUG[S]) in order to get the feeling you wanted than you did when you first started using it?

→ IF YES: How much more?

→ IF NO: What about finding that when you used the same amount, it had much less effect than before?

_____

_____

_____

In the past year, have you had any withdrawal symptoms, that is, felt sick when you cut down or stopped using (DRUG[S])?

→ IF YES: What symptoms did you have? REFER TO LIST OF WITHDRAWAL SYMPTOMS ON H.18

→ IF NO: After not using (DRUG[S]) for a few hours or more, did you sometimes use it to keep yourself from getting sick with (WITHDRAWAL SYMPTOMS)?

_____

_____

_____

(1) tolerance, as defined by either of the following: (a) a need for markedly increased amounts of the substance to achieve intoxication or desired effect (b) markedly diminished effect with continued use of the same amount of the substance	SED	?	1	2	3	H64
	STIM	?	1	2	3	H65
	OPI	?	1	2	3	H66
	CAN	?	1	2	3	H67
	HER	?	1	2	3	H68
	COC	?	1	2	3	H69
	HAL	?	1	2	3	H70
	INH	?	1	2	3	H71
(2) withdrawal, as manifested by either of the following: (a) the characteristic withdrawal syndrome for the substance (b) the same (or a closely related) substance is taken to relieve or avoid withdrawal symptoms	SED	?	1	2	3	H72
	STIM	?	1	2	3	H73
	OPI	?	1	2	3	H74
	CAN	?	1	2	3	H75
	HER	?	1	2	3	H76
	COC	?	1	2	3	H77
	HAL	?	1	2	3	H78
	INH	?	1	2	3	H79

IF UNKNOWN: When did (SXS CODED "3" ABOVE) occur? (Did they all happen around the same time?)	AT LEAST THREE DEPENDENCE ITEMS CODED "3" AND ITEMS OCCURRED WITHIN THE SAME TWELVE MONTH PERIOD	SED	1	3	H80
		STIM	1	3	H81
		OPI	1	3	H82
		CAN	1	3	H83
		HER	1	3	H84
		COC	1	3	H85
		HAL	1	3	H86
		INH	1	3	H87

Substance Dependence

NON-ALCOHOL SUBSTANCE ABUSE PAST YEAR

Now I'd like to ask you some questions about (TIME IN THE PAST YEAR WHEN USED DRUG[S] THE MOST / YOUR USE OF DRUG[S] DURING THE PAST 12 MONTHS).

During that time...

…did you miss work or school because you were very high or very hung over? (What about doing a bad job at work or failing courses at school because you used [DRUG(S)]?)

IF NO: What about not keeping your house clean [IF CHILDREN: or not taking proper care of your children] because of using (DRUG[S])?

IF YES TO EITHER: How often? (Over what period of time?)

_____

_____

_____

…have you used (DRUG[S]) in a situation in which it might have been dangerous to be using (DRUG[S]) at all? During the past year, have you driven while you were really too high to drive?)

IF YES AND UNKNOWN: How many times? (When?)

_____

_____

_____

NON-ALCOHOL SUBSTANCE ABUSE CRITERIA

A. A maladaptive pattern of substance use leading to clinically significant impairment or distress, as manifested by one (or more) of the following occurring within a twelve month period:

(1)	recurrent substance use resulting in a failure to fulfill major role obligations at work, school, or home (e.g., repeated absences or poor work performance related to substance use; substance-related absences, suspensions, or expulsions from school; neglect of children or household)	SED	?	1	2	3	H88
		STIM	?	1	2	3	H89
		OPI	?	1	2	3	H90
		CAN	?	1	2	3	H91
		HER	?	1	2	3	H92
		COC	?	1	2	3	H93
		HAL	?	1	2	3	H94
		INH	?	1	2	3	H95
(2)	recurrent substance use in situations in which it is physically hazardous (e.g., driving an automobile or operating a machine when impaired by substance use)	SED	?	1	2	3	H96
		STIM	?	1	2	3	H97
		OPI	?	1	2	3	H98
		CAN	?	1	2	3	H99
		HER	?	1	2	3	H100
		COC	?	1	2	3	H101
		HAL	?	1	2	3	H102
		INH	?	1	2	3	H103

…has your use of (DRUG[S]) gotten you into trouble with the law? IF YES AND UNKNOWN: How often? (Over what period of time?) _____ _____ _____	(3)	recurrent substance-related legal problems (e.g., arrests for substance-related disorderly conduct)	SED	?	1	2	3	H104
			STIM	?	1	2	3	H105
			OPI	?	1	2	3	H106
			CAN	?	1	2	3	H107
			HER	?	1	2	3	H108
			COC	?	1	2	3	H109
			HAL	?	1	2	3	H110
			INH	?	1	2	3	H111
IF NOT ALREADY KNOWN: Has your use of (DRUG[S]) caused problems with other people, such as with family members, friends, or people at work? (Did you get into physical fights or bad arguments about your [DRUG(S)] use?) IF YES: Did you keep on using (DRUG[S]) anyway? (Over what period of time?) _____ _____	(4)	continued substance use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of the substance (e.g., arguments with spouse about consequences of intoxication, physical fights)	SED	?	1	2	3	H112
			STIM	?	1	2	3	H113
			OPI	?	1	2	3	H114
			CAN	?	1	2	3	H115
			HER	?	1	2	3	H116
			COC	?	1	2	3	H117
			HAL	?	1	2	3	H118
			INH	?	1	2	3	H119

SUBSTANCE ABUSE

AT LEAST ONE "A" ITEM CODED "3"

SED	1	3	H120
STIM	1	3	H121
OPI	1	3	H122
CAN	1	3	H123
HER	1	3	H124
COC	1	3	H125
HAL	1	3	H126
INH	1	3	H127

Substance Abuse

LIST OF WITHDRAWAL SYMPTOMS (FROM DSM-IV CRITERIA)

Listed below are the characteristic withdrawal symptoms for those classes of psychoactive substances for which a withdrawal syndrome has been identified. (NOTE: A specific withdrawal syndrome has not been identified for CANNABIS AND HALLUCINOGENS/PCP). Withdrawal symptoms may occur following the cessation of prolonged moderate or heavy use of a psychoactive substance or a reduction in the amount used.

SEDATIVES, HYPNOTICS, AND ANXIOLYTICS:

Two (or more) of the following, developing within several hours to a few days after cessation (or reduction) of sedative, hypnotic, or anxiolytic use, which has been heavy and prolonged:

(1) autonomic hyperactivity (e.g., sweating or pulse rate greater than 100)
(2) increased hand tremor
(3) insomnia
(4) nausea or vomiting
(5) transient visual, tactile, or auditory hallucinations or illusions
(6) psychomotor agitation
(7) anxiety
(8) grand mal seizures

STIMULANTS/COCAINE

Dysphoric mood AND two (or more) of the following physiological changes, developing within a few hours to several days after cessation (or reduction of substance use which has been heavy and prolonged):

(1) fatigue
(2) vivid, unpleasant dreams
(3) insomnia or hypersomnia
(4) increased appetite
(5) psychomotor retardation or agitation

OPIOIDS:

Three (or more) of the following, developing within minutes to several days after cessation (or reduction) of opioid use which has been heavy and prolonged (several weeks or longer) or after administration of an opioid antagonist (after a period of opioid use):

(1) dysphoric mood
(2) nausea or vomiting
(3) muscle aches
(4) lacrimation or rhinorrhea
(5) pupillary dilation, piloerection, or sweating
(6) diarrhea
(7) yawning
(8) fever
(9) insomnia

J. ADJUSTMENT DISORDER

IF THERE IS A DISTURBANCE IN THE PAST YEAR AND IT DOES NOT MEET THE CRITERIA FOR ANOTHER AXIS I DSM-IV DISORDER, CIRCLE 3 AND CONTINUE. OTHERWISE, CIRCLE 1 AND GO TO ***END OF INTERVIEW*** ON PAGE K.1.

INFORMATION OBTAINED FROM OVERVIEW OF PRESENT ILLNESS WILL USUALLY BE SUFFICIENT TO RATE THE CRITERIA.

DISTURBANCE IN PAST YEAR THAT DOES NOT MEET CRITERIA FOR DSM DISORDER

NO	YES
1	3

J1

ADJUSTMENT DISORDER CRITERIA

IF UNKNOWN: Did anything happen to you just before (ONSET OF CURRENT DISTURBANCE)?

IF YES: Do you think that [STRESSOR] had anything to do with your getting [SYMPTOMS]?

DESCRIBE:

_____

_____

A. The development of emotional or behavioral symptoms in response to an identifiable stressor(s) occurring within three months of the onset of the stressor(s).

? 1 2 3 — J2

GO TO ***END OF INTERVIEW***, K.1

(What effect has [SYMPTOMS] had on you and your ability to do things?) (How upset were you?) (Has it made it hard for you to do your work or be with friends?)

_____

_____

_____

_____

B. These symptoms or behaviors are clinically significant as evidenced by either of the following:

(1) marked distress that is in excess of what would be expected from exposure to the stressor

(2) significant impairment in social or occupational (academic) functioning

? 1 2 3 — J3

GO TO ***END OF INTERVIEW***, K.1

(Have you had this kind of reaction many times before?)

(Were you having these [SYMPTOMS] even before [STRESSOR] happened?)

_____

_____

C. The stress-related disturbance does not meet the criteria for another specific Axis I disorder and is not merely an exacerbation of a preexisting Axis I or Axis II disorder.

? 1 3 — J4

GO TO ***END OF INTERVIEW***, K.1

IF UNKNOWN: Did someone close to you die just before (ONSET OF CURRENT DISTURBANCE)?

_____

_____

(How long has it been now since [STRESSOR AND COMPLICATIONS ARISING FROM THE STRESSOR] was over?)

_____

_____

D. The symptoms do not represent Bereavement. ? 1 3 J5

GO TO ***END OF INTERVIEW***, K.1

E. Once the stressor (or its consequences) has terminated, the symptoms do not persist for more than an additional 6 months. ? 1 2 3 J6

GO TO ***END OF INTERVIEW***, K.1

ADJUSTMENT DISORDER CRITERIA A, B, C, D, AND E ARE CODED "3" 1 3 J7

GO TO ***END OF INTERVIEW***, K.1

ADJUST-MENT DISORDER

CODE SUBTYPE BASED ON PREDOMINANT SYMPTOMS J8

1 WITH DEPRESSED MOOD
(e.g., depressed mood, tearfulness, feelings of hopelessness)

2 WITH ANXIETY
(e.g., nervousness, worry, jitteriness, or in children, fears of separation from major attachment figures)

3 WITH MIXED ANXIETY AND DEPRESSED MOOD
(e.g., a combination of depression and anxiety)

4 WITH DISTURBANCE OF CONDUCT
(a disturbance in conduct in which there is a violation of the rights of others or of major age-appropriate societal norms or rules, e.g., truancy, vandalism, reckless driving, fighting, defaulting on legal responsibilities)

5 WITH MIXED DISTURBANCE OF EMOTIONS AND CONDUCT (e.g., depression and disturbance of conduct)

6 UNSPECIFIED
(e.g., physical complaints, social withdrawal, or work or academic inhibition)

This page has been intentionally left blank.

BEFORE YOU END THIS ASSESSMENT, REVIEW THE INFORMATION YOU HAVE ABOUT THE RESPONDENT'S PAST YEAR SYMPTOMS AND FUNCTIONING. IN ORDER TO ACCURATELY ASSIGN A GAF SCORE ON THE NEXT PAGE, YOU NEED TO UNDERSTAND THE EXTENT TO WHICH MENTAL HEALTH/ILLNESS HAS:

- IMPAIRED/INHIBITED THE RESPONDENT'S ABILITY TO MAINTAIN A HOME, CARE FOR CHILDREN;
- IMPAIRED/INHIBITED THE RESPONDENT'S ABILITY TO FUNCTION AT WORK AND OR SCHOOL;
- IMPAIRED/IMHIBITED THE RESPONDENT'S ABILITY TO TAKE CARE OF HIM/HERSELF WITH REGARD TO PERSONAL HYGIENE AND SAFETY;
- IMPAIRED/INHIBITED THE RESPONDENT'S ABILITY TO MAINTAIN FRIENDSHIPS AND POSITIVE RELATIONSHIPS WITH FAMILY MEMBERS;
- MADE THE RESPONDENT A DANGER TO HIM/HERSELF OR OTHERS

QUERY ANY UNKNOWN DIMENSIONS OF THE RESPONDENT'S PAST YEAR SYMPTOMATOLOGY AND FUNCTIONING, AND ASSIGN A GAF SCORE ON THE NEXT PAGE.

DSM-IV Axis V: Global Assessment of Functioning Scale

Consider psychological, social, and occupational functioning on a hypothetical continuum of mental health-illness. Rate the respondent's period of worst functioning in the past year. Do not include impairment in functioning due to physical (or environmental) limitations.

CODE (**Note:** Use intermediate codes when appropriate, e.g., 45, 68, 72). ___ ___ ___ EOI1

100 | 91 **Superior functioning in a wide range of activities, life's problems never seem to get out of hand, is sought out by others because of his many positive qualities. No symptoms.**

90 | 81 **Absent or minimal symptoms** (e.g., mild anxiety before an exam); **good functioning in all areas, interested and involved in a wide range of activities, socially effective, generally satisfied with life, no more than everyday problems or concerns** (e.g., an occasional argument with family members).

80 | 71 **If symptoms are present, they are transient and expectable reactions to psychosocial stressors** (e.g., difficulty concentrating after family argument); **no more than slight impairment in social, occupational, or school functioning** (e.g., temporarily falling behind in school work).

70 | 61 **Some mild symptoms** (e.g., depressed mood and mild insomnia) **OR some difficulty in social, occupational, or school functioning** (e.g., occasional truancy, or theft within the household), **but generally functioning pretty well, has some meaningful interpersonal relationships.**

60 | 51 **Moderate symptoms** (e.g., flat affect and circumstantial speech, occasional panic attacks) **OR moderate difficulty in social, occupational, or school functioning** (e.g., few friends, conflicts with peers or coworkers).

50 | 41 **Serious symptoms** (e.g., suicidal ideation, severe obsessional rituals, frequent shoplifting) **OR any serious impairment in social, occupational, or school functioning** (e.g., no friends, unable to keep a job).

40 | 31 **Some impairment in reality testing or communication** (e.g., speech is at times illogical, obscure, or irrelevant) **OR major impairment in several areas, such as work or school, family relations, judgment, thinking, or mood** (e.g., depressed man avoids friends, neglects family, and is unable to work; child frequently beats up younger children, is defiant at home, and is failing at school).

30 | 21 **Behavior is considerably influenced by delusions or hallucinations OR serious impairment in communication or judgment** (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) **OR inability to function in almost all areas** (e.g., stays in bed all day; no job, home, or friends).

20 | 11 **Some danger of hurting self or others** (e.g., suicide attempts without clear expectation of death, frequently violent, manic excitement) **OR occasionally fails to maintain minimal personal hygiene** (e.g., smears feces) **OR gross impairment in communication (e.g., largely incoherent or mute).**

10 | 1 **Persistent danger of severely hurting self or others** (e.g., recurrent violence) **OR persistent inability to maintain minimal personal hygiene OR serious suicidal act with clear expectation of death.**

0 Inadequate information.

That was my last question. Thank you for your time and cooperation in completing this interview.

Sometimes the personal issues we've discussed cause people to become upset and in need of speaking with a counselor. If you are feeling upset or disturbed by the personal issues we have discussed in this interview and would like to talk with someone about your feelings, we suggest you call your doctor, counselor, or other treatment provider if you are currently under someone's care. If not, there is also a national lifeline number you can call. This number is on the receipt for the $30 you received for this interview from the interviewer who met with you earlier. Do you still have that receipt?

IF NO: We would like to give you the hotline number for the National Lifeline Network where counselors are available to talk at any time of the day or night. They can also give you information about (additional) mental health services in your area. Their toll-free number is 1-800-273-8255.

IF YES: OK. Please know that counselors at the National Lifeline are available to talk at any time of the day or night. They can also give you information about mental health services in your area if you request this information.

Do you have any additional questions you'd like to ask me before we end our call?

Thank you again for your time, and have a good (day/afternoon/evening).

Interview End Time: ___ ___ : ___ ___ *AM/PM*

INTERVIEWER DEBRIEFING SECTION

Distressed Respondent Protocol

	No	Yes	
Was the Distressed Respondent Protocol used?	1	3	IDS1
Specify problems:			IDS2

_____

_____

_____

Cognitive Impairment Screener

	No	Yes	
Was the Short Blessed Scale used?	1	3	IDS3 IF IDS3 = 1, SKIP IDS4 and IDS4a
Specify problems:			IDS4

_____

_____

_____

Indicate score on the Short Blessed _____ (0-28) IDS4a

Stressful Life Circumstances

Were there significant problems in these areas?	No	Yes	
Problems with primary support group	1	3	IDS5
Problems related to social environment	1	3	IDS6
Educational problems	1	3	IDS7
Occupational problems	1	3	IDS8
Housing problems	1	3	IDS9
Economic problems	1	3	IDS10
Problems with access to health care services	1	3	IDS11
Problems related to interaction with the legal system/crime	1	3	IDS12
Life-threatening Illness – self	1	3	IDS13
Life-threatening illness – partner, spouse, family member	1	3	IDS14
Other psychosocial and environmental problems	1	3	IDS15

Comprehension Rating

Estimate the respondent's understanding of the interview:	***Circle response***	IDS16
No difficulty -- No language or comprehension problem	1	
Just a little difficulty – almost no language or comprehension problems	2	
A fair amount of difficulty - some language or comprehension problems	3	
A lot of difficulty – considerable language or comprehension problems	4	
Extreme problems with language or comprehension problems	5	

Specify problems: IDS17

_____

_____

_____

Cooperation Rating

Rate how cooperative the respondent was during the interview	***Circle response***	IDS18
Very Cooperative	1	
Fairly Cooperative	2	
Not Very Cooperative	3	
Uncooperative	4	
Openly Hostile	5	

Specify problems: IDS19

_____

_____

_____

Privacy Rating

Indicate on a scale of 1 through 5 how private the interview was:	***Circle response***	IDS20
Completely Private – No one who could overhear any part of the interview appeared present	1	
Minor Distractions – Other person(s) seemed present or listening for less than 1/3 of the time	2	
Moderate Distractions – Others seemed to present about 1/3 of the time	3	
Severe Distractions - Interruptions of Privacy More Than Half the Time	4	
Constant Presence of Other Person(s)	5	

Specify problems: IDS21

_____

_____

_____

Global Validity Rating

Rate the overall validity of the interview	***Circle response***	IDS22
Excellent, no reason to suspect invalid responses	1	
Good, factors present that may adversely affect validity	2	
Fair, factors present that definitely reduce validity	3	
Poor, substantially reduced validity	4	
Invalid responses, severely impaired mental status or possible deliberate "faking bad" or "faking good"	5	

Specify problems: IDS23

_____

_____

_____

Potential Disorders Not Assessed

	No	Yes	
Were there any disorders not assessed that would need to be ruled out?	1	3	IDS24
			IF IDS24 = 1, SKIP IDS24a and IDS25
Rule-out disorder present	1 Rule-out Other Axis I Disorder (not assessed in study) 2 Rule-out Axis II Disorder – Personality Disorder (not assessed in study) 3 Rule-out Axis II Disorder – Other (e.g. Developmental Disability) (not assessed in study) 4 Rule-out Axis I Disorder assessed but missed (due to CI or R error)		IDS24a

Specify disorders implicated: IDS25

_____

_____

_____

CLINICAL SUPERVISOR'S RATINGS

CS: Global Validity Rating

Rate the overall validity of the interview	***Circle response***	IDS26
Excellent, no reason to suspect invalid responses	1	
Good, factors present that may adversely affect validity	2	
Fair, factors present that definitely reduce validity	3	
Poor, substantially reduced validity	4	
Invalid responses, severely impaired mental status or possible deliberate "faking bad" or "faking good"	5	

Specify problems: IDS27

_____

_____

_____

CS: Potential Disorders Not Assessed

	No	**Yes**	
Were there any disorders not assessed that would need to be ruled out?	1	3	IDS28

IF IDS28 = 1, SKIP IDS28a and IDS29

Rule-out disorder present IDS28a

1 Rule-out Other Axis I Disorder (not assessed in study)
2 Rule-out Axis II Disorder – Personality Disorder (not assessed in study)
3 Rule-out Axis II Disorder – Other (e.g. Developmental Disability) (not assessed in study)
4 Rule-out Axis I Disorder assessed but missed (due to CI or R error)

Specify disorders implicated: IDS29

_____

_____

_____

This page has been intentionally left blank.

GMC/SUBSTANCE CAUSING MOOD SYMPTOMS

MOOD DISORDER DUE TO A GENERAL MEDICAL CONDITION

MOOD DISORDER DUE TO A GENERAL MEDICAL CONDITION CRITERIA

IF SYMPTOMS NOT TEMPORALLY ASSOCIATED WITH A GENERAL MEDICAL CONDITION, CHECK HERE ___ AND GO TO ***SUBSTANCE-INDUCED MOOD DISORDER,*** MDGS.3.

CODE BASED ON INFORMATION ALREADY OBTAINED

A. A prominent and persistent disturbance in mood predominates in the clinical picture and is characterized by either (or both) of the following:

(1) depressed mood or markedly diminished interest or pleasure in all, or almost all, activities ? 1 2 3

(2) elevated, expansive, or irritable mood ? 1 2 3

Do you think your (MOOD SXS) were in any way related to your (COMORBID GENERAL MEDICAL CONDITION)?

IF YES: Tell me how.

_____

(Did the [MOOD SXS] start or get much worse only after [COMORBID GENERAL MEDICAL CONDITION] began?)

IF YES AND GMC HAS RESOLVED: Did the (MOOD SXS) get better once the (COMORBID GENERAL MEDICAL CONDITION) got better?

B./C. There is evidence from the history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition and the disturbance is not better accounted for by another mental disorder (e.g., Adjustment Disorder With Depressed Mood, in response to the stress of having a general medical condition). ? 1 2 3

GO TO ***SUB-STANCE INDUCED*,** MDGS.3

THE FOLLOWING FACTORS SHOULD BE CONSIDERED AND SUPPORT THE CONCLUSION THAT THE GMC IS ETIOLOGIC TO THE MOOD SYMPTOMS:

1) THERE IS EVIDENCE FROM THE LITERATURE OF A WELL-ESTABLISHED ASSOCIATION BETWEEN THE GMC AND MOOD SYMPTOMS.

2) THERE IS A CLOSE TEMPORAL RELATIONSHIP BETWEEN THE COURSE OF THE MOOD SYMPTOMS AND THE COURSE OF THE GENERAL MEDICAL CONDITION.

3) THE MOOD SYMPTOMS ARE CHARACTERIZED BY UNUSUAL PRESENTING FEATURES (E.G., LATE AGE AT ONSET)

4) THERE ARE NO ALTERNATIVE EXPLANATIONS (E.G., MOOD SYMPTOMS AS A PSYCHOLOGICAL REACTION TO THE GMC).

IF UNCLEAR: How much did (MOOD SYMPTOMS) interfere with your life?

E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning ? 1 2 3

GO TO ***SUB-STANCE INDUCED*,** MDGS.3

NOTE: THE D CRITERION (DELIRIUM R/O) HAS BEEN OMITTED.

MOOD DISORDER DUE TO GMC CRITERIA A, B/C, AND E CODED "3" 1 3

MOOD DIS-ORDER DUE TO A GMC

CHECK HERE ___ IF CURRENT IN PAST MONTH

Indicate which type of symptom presentation predominates:
1 - With Major Depressive-like episode
2 - With Depressive Features
(if predominant mood is depressed but the full criteria are not met for a Major depressive episode)
3 - With Manic Features
4 - With Mixed Features

CONTINUE ON NEXT PAGE

SUBSTANCE-INDUCED MOOD DISORDER

SUBSTANCE-INDUCED MOOD DISORDER CRITERIA

EPISODE BEING EVALUATED:	
Past Year MDE	A.5
Lifetime MDE	A.11
Past Year Manic	A.17
Lifetime Manic	A.22
Dysthymic	A.29

IF SYMPTOMS NOT TEMPORALLY ASSOCIATED WITH SUBSTANCE, CHECK HERE ___ AND RETURN TO EPISODE BEING EVALUATED.

CODE BASED ON INFORMATION ALREADY OBTAINED.

A. A prominent and persistent disturbance in mood predominates in the clinical picture and is characterized by one (or both) of the following:

(1) depressed mood or markedly diminished interest or pleasure in all, or almost all, activities ? 1 2 3

(2) elevated, expansive or irritable mood ? 1 2 3

IF NOT KNOWN: When did the (MOOD SYMPTOMS) begin? Were you already using (SUBSTANCE) or had you just stopped or cut down your use?

B. There is evidence from the history, physical examination or laboratory findings that either (1) the symptoms in A developed during or within a month of substance intoxication or withdrawal, or (2) medication use is etiologically related to the disturbance ? 1 2 3

NOT SUBSTANCE INDUCED RETURN TO EPISODE BEING EVALUATED

Do you think your (MOOD SXS) are in any way related to your (SUBSTANCE USE)?

IF YES: Tell me how.

ASK ANY OF THE FOLLOWING QUESTIONS AS NEEDED TO RULE OUT A NON-SUBSANCE-INDUCED ETIOLOGY

C. The disturbance is not better accounted for by a Mood Disorder that is not substance-induced. Evidence that the symptoms are better accounted for by a Mood Disorder that is not substance-induced might include: ? 1 2 3

NOT SUBSTANCE INDUCED RETURN TO EPISODE BEING EVALUATED

IF UNKNOWN: Which came first, the (SUBSTANCE USE) or the (MOOD SYMPTOMS)?

1) the mood symptoms precede the onset of the Substance Abuse or Dependence (or medication use)

IF UNKNOWN: Have you had a period of time when you stopped using (SUBSTANCE)?

IF YES: After you stopped using (SUBSTANCE) did the (MOOD SXS) get better?

2) the mood symptoms persist for a substantial period of time (e.g., about a month) after the cessation of acute withdrawal or severe intoxication

IF UNKNOWN: How much of (SUBSTANCE) were you using when you began to have (MOOD SYMPTOMS)?

3) the mood symptoms are substantially in excess of what would be expected given the type, duration or amount of the substance used

IF UNKNOWN: Have you had any other episodes of (MOOD SYMPTOMS)?

IF YES: How many? Were you using (SUBSTANCES) at those times?

4) there is evidence suggesting the existence of an independent non-substance-induced Mood Disorder (e.g. , a history of recurrent Major Depressive Episodes)

IF UNKNOWN: How much did (MOOD SYMPTOMS) interfere with your life?

E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning. ? 1 2 3

RETURN TO EPISODE BEING EVALUATED

NOTE: THE D CRITERION (DELIRIUM R/O) HAS BEEN OMITTED.

SUBSTANCE-INDUCED MOOD DISORDER CRITERIA A, B, C, AND E ARE CODED "3"

1 3

SUBSTANCE-INDUCED MOOD DISORDER

CHECK HERE ___ IF CURRENT IN PAST MONTH

Indicate which type of symptom presentation predominates:
1 – With Depressive Features
2 – With Manic Features
3 – With Mixed Features

Indicate context of development of mood symptoms:
1 – With Onset During Intoxication
2 – With Onset During Withdrawal

RETURN TO EPISODE BEING EVALUATED

GMC/SUBSTANCE AS ETIOLOGY FOR ANXIETY SYMPTOMS

ANXIETY DISORDER DUE TO A GENERAL MEDICAL CONDITION

ANXIETY DISORDER DUE TO A GENERAL MEDICAL CONDITION CRITERIA

IF SYMPTOMS NOT TEMPORALLY ASSOCIATED WITH A GENERAL MEDICAL CONDITON CHECK HERE ___ AND GO TO ***SUBSTANCE-INDUCED ANXIETY DISORDER,*** ADGS.3

CODE BASED ON INFORMATION ALREADY OBTAINED

A. Prominent anxiety, panic attacks, obsessions or compulsions predominate in the clinical picture. ? 1 3

Did the (ANXIETY SYMPTOMS) start or get much worse only after (GMC) began?

IF GMC HAS RESOLVED: Did the (ANXIETY SYMPTOMS) get better once the (GMC) got better?

_____

_____

B/C. There is evidence from this history, physical examination, or laboratory findings that the disturbance is the direct physiological consequence of a general medical condition and the disturbance is not better accounted for by another mental disorder (e.g., adjustment disorder With Anxiety), in which the stressor is a serious general medical condition). ? 1 2 3

GO TO ***SUBSTANCE INDUCED*** ADGS.3

THE FOLLOWING FACTORS SHOULD BE CONSIDERED AND SUPPORT THE CONCLUSION THAT THE GMC IS ETIOLOGIC TO THE ANXIETY SYMPTOMS.

1) THERE IS EVIDENCE FROM THE LITERATURE OF A WELL-ESTABLISHED ASSOCIATION BETWEEN THE GMC AND ANXIETY SYMPTOMS.

2) THERE IS A CLOSE TEMPORAL RELATIONSHIP BETWEEN THE COURSE OF THE ANXIETY SYMPTOMS AND THE COURSE OF THE GENERAL MEDICAL CONDITION.

3) THE ANXIETY SYMPTOMS ARE CHARACTERIZED BY UNUSUAL PRESENTING FEATURES (E.G., LATE AGE AT ONSET)

4) THE ABSENCE OF ALTERNATIVE EXPLANATIONS (E.G., ANXIETY SYMPTOMS AS A PSYCHOLOGICAL REACTION TO THE GMC).

IF UNCLEAR: How much did (ANXIETY SYMPTOMS) interfere with your life?

(Has it made it hard for you to do your work or be with your friends?)

_____

_____

E. The disturbance causes clinically significant distress or impairment in social, occupational or other important areas of functioning. ? 1 2 3

GO TO ***SUBSTANCE INDUCED*** ADGS.3

NOTE: THE D CRITERION (DELIRIUM R/O) HAS BEEN OMITTED.

ANXIETY DISORDER DUE TO GMC CRITERIA A, B/C, AND E CODED "3" 1 3

ANXIETY DISORDER DUE TO A GMC

CHECK HERE ___ IF CURRENT IN PAST MONTH

Indicate which type of symptom presentation predominates:
1 - With Generalized Anxiety
2 - With Panic attacks
3 - With Obsessive-Compulsive symptoms

CONTINUE ON NEXT PAGE

SUBSTANCE-INDUCED ANXIETY DISORDER

SUBSTANCE-INDUCED ANXIETY DISORDER CRITERIA

EPISODE BEING EVALUATED:	
Panic	E.12
AWOPD	E.17
Social Phobia	E.21
OCD	E.31
GAD	E.36

IF SYMPTOMS NOT TEMPORALLY ASSOCIATED WITH SUBSTANCE USE, CHECK HERE ___ AND RETURN TO DISORDER BEING EVALUATED.

CODE BASED ON INFORMATION ALREADY OBTAINED

A. Prominent anxiety, panic attacks, obsessions or compulsions predominate in the clinical picture. ? 1 2 3

IF NOT KNOWN: When did the (ANXIETY SYMPTOMS) begin? Were you already using (SUBSTANCE) or had you just stopped or cut down your use?

_____

_____

B. There is evidence from the history, physical examination, or laboratory findings that either: (1) the symptoms in A developed during, or within a month of, substance intoxication or withdrawal, or (2) medication use is etiologically related to the disturbance ? 1 2 3

NOT SUBSTANCE INDUCED

RETURN TO DISORDER BEING EVALUATED

ASK ANY OF THE FOLLOWING QUESTIONS AS NEEDED TO RULE OUT A NON-SUBSTANCE-INDUCED ETIOLOGY:

C. The disturbance is NOT better accounted for by an Anxiety Disorder that is not substance-induced. ? 1 3

NOT SUBSTANCE INDUCED

RETURN TO DISORDER BEING EVALUATED

<u>Guidelines for Primary Anxiety:</u>
Evidence that the symptoms are better accounted for by a primary (i.e., non-substance-induced) Anxiety Disorder may include any (or all) of the following:

IF UNKNOWN: Which came first, the (SUBSTANCE USE) or the (ANXIETY SYMPTOMS)?

IF UNKNOWN: Have you had a period of time when you stopped using (SUBSTANCE)?

IF YES: After you stopped using (SUBSTANCE) did the (ANXIETY SYMPTOMS) get better or did they continue?

_____

_____

(1) the anxiety symptoms precede the onset of the Substance Abuse or Dependence (or medication use)

(2) the anxiety symptoms persist for a substantial period of time (e.g., about a month) after the cessation of acute withdrawal or severe intoxication

IF UNKNOWN: How much (SUBSTANCE) were you using when you began to have (ANXIETY SYMPTOMS)?

_____

_____

(3) the anxiety symptoms are substantially in excess of what would be expected given the character, duration, or amount of the substance used

IF UNKNOWN: Have you had any other episodes of (ANXIETY SYMPTOMS)?

IF YES: How many? Were you using (SUBSTANCES) at those times?

_____

_____

(4) there is evidence suggesting the existence of an independent, non-substance-induced Anxiety Disorder (e.g., a history of recurrent non-substance-related panic attacks)

IF UNKNOWN: How much did (ANXIETY SYMPTOMS) interfere with your life?

(Has it made it hard for you to do your work or be with your friends?)

_____

_____

E. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

? 1 2 3

RETURN TO DISORDER BEING EVALUATED

NOTE: THE D CRITERION (DELIRIUM R/O) HAS BEEN OMITTED.

SUBSTANCE-INDUCED ANXIETY DISORDER CRITERIA A, B, C, AND E ARE CODED "3"
1
3
SUBSTANCE-INDUCED ANXIETY DISORDER
CHECK HERE ___ IF CURRENT IN PAST MONTH
Indicate which type of symptom presentation predominates:
0 - Unspecified
1 - With Generalized Anxiety
2 - With Panic Attacks
3 - With Obsessive-Compulsive symptoms
4 - With Phobic Symptoms
Indicate context of development of anxiety symptoms:
1- With Onset During Intoxication
2- With Onset During Withdrawal
RETURN TO EPISODE BEING EVALUATED

This page has been intentionally left blank.

Specific Guidelines

If respondents report any of the issues listed below during any interactions with the recruiter or clinical interviewer, including before, during, or after a telephone screening or interview, the staff member will immediately refer to the scenario chart below and follow the instructions provided. Details of all incidents will be documented on the case management system and reported to project management staff immediately.

- Has had **any suicidal thoughts in the past two weeks (p. A.3)**, including
  - passive suicidal thoughts (i.e., thoughts or wishes about his/her death **in the absence of** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, or intention of dying or attempting suicide) **[SCENARIO 1]** or
  - active suicidal thoughts (i.e., thoughts or wishes about his/her death **combined with** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, the intention of dying or attempting suicide, the means to carry out that plan, and will not promise you that s/he will not hurt him/herself) **[SCENARIO 2]**

- Has had **any homicidal thoughts in the past two weeks**, including
  - passive homicidal thoughts (i.e., thoughts or wishes about seriously harming someone else **in the absence of** thoughts about specific ways in which s/he could seriously harm another person, plans for how s/he could seriously harm another person, intentions of seriously harming another person) **[SCENARIO 3]** or
  - active homicidal thoughts (i.e., thoughts or wishes about seriously harming someone else **combined with** thoughts about specific ways s/he could seriously harm another person, plans for how s/he could seriously harm another person, the intention of seriously harming another person, and the means to carry out that plan) **[SCENARIO 4]**

Scenario Chart		
Scenario Number	**Individual at Risk of Harm**	**Imminent Danger?**
1	Self	No
2	Self	Possible / Yes
3	Other(s)	No
4	Other(s)	Possible / Yes
5	No risk of harm; respondent is agitated or upset	No

Scenario Number	Individual at Risk of Harm	Imminent Danger?
1	Self	No

STEPS

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about your death or dying.

B. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about death and dying. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

C. WHEN CALL IS COMPLETED, CALL IF YOU HAVE QUESTIONS OR WOULD LIKE TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

Scenario Number	Individual at Risk of Harm	Imminent Danger?
2	Self	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you are thinking about harming yourself. Can you promise me that you will not harm yourself today?

IF NO: GO TO STEP B.

IF YES: Do you have a doctor, counselor, or other professional you can talk to about how you are feeling?

> **IF YES:** I strongly suggest that you contact this person so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about death and dying. Would you be willing to do that?
>
> > **IF YES:** Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. I strongly suggest you contact counselors at Lifeline. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**
>
> **IF NO:** I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. I strongly suggest that we contact emergency care services in your area, such as a crisis center or nearby hospital. I am going to look-up that number. Can you remain on the line while I do that? It may take a few minutes.

> **IF NO:** Okay, if I don't connect you with the local emergency care provider, then I will need to call the provider myself to see if they can send someone to you who can provide the care you need in order to keep you safe. I'll call you back to let you know what I find out.

C. FIND THE NEAREST EMERGENCY PSYCHIATRIC SERVICES USING THE SAMHSA WEBSITE (http://mentalhealth.samhsa.gov/databases/). SEARCH FOR <u>INPATIENT</u> MH TREATMENT USING THE R'S CURRENT ZIP CODE.

D. CALL THEIR LOCAL INPATIENT PSYCHIATRIC CARE FACILITY OR CRISIS CENTER AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming (himself/herself) and I am concerned about (his/her) safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

IF ASKED FOR NSDUH OVERVIEW: This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (his/her) well-being, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

E. GIVE R FIRST NAME, TELEPHONE NUMBER, AND ADDRESS (IF KNOWN) TO LOCAL EMERGENCY CARE REPRESENTATIVE. IF THEY ARE UNABLE TO PROVIDE SERVICES THAT ENSURE THE R'S SAFETY, SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

F. IF R <u>NOT</u> ON THE OTHER LINE, END CALL WITH THE EMERGENCY CARE PROVIDER OR LOCAL 911 DISPATCHER AND ATTEMPT TO CONTACT R AGAIN WITH AN UPDATE.

IF R ON THE OTHER LINE, CONNECT R TO EMERGENCY CARE REPRESENTATIVE OR LOCAL 911 DISPATCHER AND STAY ON THE LINE; IF YOU HANG-UP, THEIR CONNECTION WILL ALSO END.

G. YOU MAY STAY ON THE LINE TO WAIT FOR THE RESCUE TEAM TO ARRIVE. IF SO, DO NOT CONTINUE THE INTERVIEW. KEEP THE DISCUSSION LIGHT AND AVOID EMOTIONAL TOPICS. DEMONSTRATE EMPATHIC LISTENING BUT REFRAIN FROM COUNSELING OR PRACTICING PSYCHOLOGY.

H. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT OR TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Scenario Number	Individual at Risk of Harm	Imminent Danger?
3	Other(s)	No

STEPS

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about seriously harming someone else. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about seriously harming someone else. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF DIRECTED BY ONE OF THEM, FOLLOW SCENARIO 4 FOR POSSIBLE IMMINENT DANGER TO OTHERS. FILL OUT ONLINE INCIDENT REPORT.

Scenario Number	Individual at Risk of Harm	Imminent Danger?
4	Other(s)	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND END CALL.

B. SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

C. CALL THEIR LOCAL 911, AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming another individual. I am concerned about this individual's safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

> **IF ASKED FOR NSDUH OVERVIEW:** This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (him/her) harming someone else, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

D. GIVE R FIRST NAME, TELEPHONE NUMBER, ADDRESS (IF KNOWN), AND VICTIM'S IDENTIFYING INFORMATION TO LOCAL 911 DISPATCHER. END CALL.

E. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT OR TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Scenario Number	Individual at Risk of Harm	Imminent Danger?
5	No risk of harm; respondent is agitated or upset	No

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: I know these questions are very personal, and they seem to be upsetting you. Do you have a doctor or someone you can talk to about how you are feeling?

IF YES: I suggest that you call that individual immediately so that she or he can help you talk about and work through how you are feeling. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline is a 24-hour hotline that you could call to discuss this with a counselor. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away or call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL IF YOU HAVE ANY QUESTIONS OR NEED TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

This page has been intentionally left blank.

SHORT BLESSED SCALE EXAM

THE SHORT BLESSED SCALE IS TO BE COMPLETED AT ANY POINT DURING THE INTERVIEW IF THE RESPONDENT APPEARS TO BE COGNITIVELY IMPAIRED.

ERROR SCORES

SB-1. What year is it now? _____

CIRCLE 4 FOR ANY ERROR 0 4

SB-2. What month is it now? _____

CIRCLE 3 FOR ANY ERROR 0 3

Please repeat this phrase after me: John Brown, 42 Market Street, Chicago.

NO SCORE – FOR ITEM SB-6.

SB-3. About what time is it? _____

CIRCLE 3 FOR ANY ERROR 0 3

SB-4. Please count backwards from 20 to 1.
[20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1]

2 PER ERROR 0 2 4

SB-5. Please say the months of the year in reverse order.
[DEC, NOV, OCT, SEP, AUG, JUL, JUN, MAY, APR, MAR, FEB, JAN]

2 PER ERROR 0 2 4

SB-6. Please repeat the phrase I asked you to repeat before.
[JOHN BROWN / 42 MARKET STREET / CHICAGO]

2 PER ERROR 0 2 4 6 8 10

TOTAL NUMBER OF ERRORS IN SB-1 TO SB-6: ..... _____

IF THE TOTAL NUMBER OF ERRORS IS GREATER THAN 10, TERMINATE THE INTERVIEW.

This page has been intentionally left blank.

Appendix D: Mental Health Surveillance Study Clinical Interviewer Handbook

Addendum 1: Skype User Guide……..S-1
Addendum 2: Distressed Respondent Protocol Booklet……..DRP-1

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study
Clinical Interviewer Handbook

Contract No. HHSS283200800004C

Project No. 0211838

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

June 2011

Mental Health Study Contacts and Information

Data Collection Managers: For assignments, case management, reporting, status updates, CMS and administrative questions.

Data Collection Mgr:		
Work Phone:		
E-mail:		
Cell Phone:		

Note: If your assigned Data Collection Manager is unavailable and your situation requires immediate assistance, please contact the other Data Collection Manager.

Clinical Supervisor: For questions on SCID administration and quality reviews.

Clinical Supervisor:			
Work Phone:			
E-mail:			
Cell Phone:			

Note: If your usual Clinical Supervisor is unavailable and your situation requires immediate assistance, please contact another supervisor listed above.

Distressed Respondent Situations: Refer to protocol in Chapter 5.

Call:	**Phone Number:**

Placing a call using the Calling Card:

1. Dial *67 (optional)
2. Dial the toll-free number (back of card): _____
3. Enter the card number (front or back of card) : _____
4. Enter 0 + Area Code + Telephone number you wish to call

Technical Support:

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study Clinical Interviewer Handbook

Contract No. HHSS283200800004C

Project No. 0211838

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

June 2011

Research Triangle Institute

MISSION

To improve the human condition by turning knowledge into practice.

VISION

To be the world's leading independent research organization, recognized for solving critical social and scientific problems.

VALUES

Integrity - We perform with the highest ethical standards of individual and group honesty. We communicate openly and realistically with each other and with our clients.

Excellence - We strive to deliver results with exceptional quality and value.

Innovation - We encourage multidisciplinary collaboration, creativity and independent thinking in everything we do.

Respect for the Individual - We treat one another fairly, with dignity and equity. We support each other to develop to our full potential.

Respect for RTI - We recognize that the strength of RTI International lies in our commitment, collectively and individually, to RTI's vision, mission, values, strategies and practices. Our commitment to the Institute is the foundation for all other organizational commitments.

Fiscal Responsibility - We operate with financial integrity and transparency. We are accountable for cost competitiveness and continuing financial responsibility.

Objectivity - Our work is independent of undue influences by political, economic, or other factors. We maintain the highest level of scientific objectivity in our work.

TABLE OF CONTENTS

1. INTRODUCTION TO THE STUDY 1-1

1.1 Research Triangle Institute (RTI) 1-1

1.2 Brief History of NSDUH 1-1

1.3 Overview of the Mental Health Surveillance Study 1-2

1.4 Study Sample 1-3

1.5 Schedule 1-3

1.6 Organization 1-4

1.7 Project Abbreviations and Terminology 1-6

1.8 Use of the Mental Health Study Handbook 1-6

2. YOUR ROLE ON THE MENTAL HEALTH STUDY 2-1

2.1 Introduction 2-1

2.2 Your Responsibilities 2-1

2.3 Professional Ethics and Respondents' Rights 2-3

2.4 Importance of Confidentiality 2-4

2.5 Protecting Both Confidentiality and Study Integrity 2-6

2.6 Materials, Supplies, and Equipment 2-6

3. EQUIPMENT USE AND CARE 3-1

3.1 Introduction 3-1

3.2 The NSDUH Laptop 3-1

3.2.1 Initial Set Up 3-1

3.2.2 Login Procedures 3-3

3.2.3 Locking the Laptop 3-3

3.2.4 Shutting Down the Laptop 3-4

3.3 Accessing the CMS 3-4

3.4 RTI E-Mail 3-5

3.5 Calling Card 3-6

3.6 Audio Files 3-7

3.6.1 Connecting and Using the MH Case Manager 3-7

3.6.2 Initial Test Audio Files 3-8

TABLE OF CONTENTS (Continued)

3.6.3 Recording the Interview 3-10
3.6.4 Uploading Audio Files to RTI 3-11
3.7 Technical Support Contact Information 3-12

4. **CONTACTING RESPONDENTS** 4-1
4.1 Introduction 4-1
4.2 Case Assignment 4-1
4.2.1 Case Identification Information 4-1
4.2.2 Case Assignment Details 4-1
4.3 Contacting the Selected Respondent 4-3
4.3.1 Preparation 4-3
4.3.2 Initial Contacts 4-4
4.3.3 Follow-Up with Respondents 4-8
4.4 Beginning the Clinical Interview 4-8
4.4.1 Informed Consent 4-8
4.4.2 Explaining the Survey and Answering Questions 4-9
4.4.3 Recording the Interview 4-9

5. **CONDUCTING THE INTERVIEW** 5-1
5.1 Introduction 5-1
5.2 Conduct During the Interview 5-1
5.3 SCID Overview 5-1
5.3.1 SCID Content 5-1
5.3.2 SCID Format 5-2
5.4 Administering the SCID 5-4
5.4.1 Asking the Questions 5-4
5.4.2 Common Pitfalls to Avoid 5-4
5.4.3 If You Feel the Respondent Is Not Providing Valid Data 5-4
5.5 Scoring the Symptoms 5-5
5.6 Concluding the Clinical Interview 5-7
5.7 Handling Unusual Situations 5-10
5.7.1 Breakoff 5-10

TABLE OF CONTENTS (Continued)

5.7.2 Refusal 5-10
5.7.3 Cognitive Impairment Issues 5-10
5.7.4 Distressed Respondents 5-12
5.7.5 Unable to Contact 5-22
5.7.6 Other Non-Response 5-22

6. **DOCUMENTING RESULTS** 6-1
6.1 Introduction 6-1
6.2 Documenting Results 6-1
6.2.1 Record of Calls 6-1
6.2.2 Status Codes 6-1
6.2.3 Recording Comments 6-2
6.2.4 Entering Records in the CMS 6-5
6.3 Disposition of Materials 6-6
6.4 Summary of the Interview Process 6-6

7. ADMINISTRATIVE PROCEDURES 7-1
7.1 Introduction 7-1
7.2 Terms of Employment 7-1
7.3 Working with your Supervisors 7-1
7.3.1 Conference Calls 7-1
7.3.2 E-mail 7-2
7.3.3 Web Updates 7-2
7.4 Electronic Production Time and Expense (ePTE) Reports 7-2
7.5 The ePTE System 7-2
7.5.1 Overview 7-2
7.5.2 Default Value Settings 7-3
7.5.3 Entry Form 7-5
7.5.4 Detail Form 7-7
7.5.5 Edit Checks in the ePTE System 7-18
7.5.6 Advanced Features 7-19
7.6 Submitting Receipts with ePTEs 7-22

TABLE OF CONTENTS (Continued)

7.7 Authorization for Expenditures 7-24
7.8 FedEx Mailing Procedures 7-24
7.8.1 Proper Handling of FedEx Packages 7-24

8. QUALITY CONTROL 8-1
8.1 Introduction 8-1
8.2 Staff Training and Evaluation 8-1
8.3 CI Certification 8-1
8.4 Supervision and Feedback during Data Collection 8-1
8.5 Ongoing Quality Improvements. 8-2

LIST OF EXHIBITS

Exhibit 1.1 Project Staff Information 1-5
Exhibit 1.2 List of Abbreviations 1-7
Exhibit 1.3 Definitions of Project Terminology 1-8

Exhibit 2.1 CI Data Collection Process 2-2
Exhibit 2.2 Confidentiality Pledge 2-5

Exhibit 4.1 Clinical Interviewer Introduction Script 4-5
Exhibit 4.2 Follow-up Study Description 4-10
Exhibit 4.3 Answering Questions about the Follow-up Interview 4-11

Exhibit 5.1 Sample SCID Page 5-3
Exhibit 5.2 Dos and Don’ts of SCID Administration 5-6
Exhibit 5.3 End of Interview Script 5-8
Exhibit 5.4 Follow-up Interview Payment Receipt 5-9
Exhibit 5.5 Short Blessed Scale Exam 5-11
Exhibit 5.6 Distressed Respondent Protocol 5-14
Exhibit 5.7 Incident Report Form 5-21

Exhibit 6.1 Pending Status Codes 6-3
Exhibit 6.2 Final Status Codes 6-4
Exhibit 6.3 SCID Cover Sheet Transmittal Record 6-7
Exhibit 6.4 Confidential Information Sheet 6-8
Exhibit 6.5 CI Case Checklist 6-9

Exhibit 7.1 ePTE Working Copy 7-10
Exhibit 7.2 Properly Completed ePTE 7-14
Exhibit 7.3 Headway Expense Report 7-23
Exhibit 7.4 Instructions for Completing a FedEx Airbill 7-25
Exhibit 7.5 Completed FedEx Airbill to RTI 7-26
Exhibit 7.6 Completed FedEx Airbill to RTI Technical Support 7-27

LIST OF APPENDICES

Appendix A NSDUH Respondent Materials A-1
Appendix B Instructions for Completing Paper PT&Es B-1
Appendix C The NSDUH Laptop C-1
Appendix D SCID Specifications D-1

1. INTRODUCTION TO THE STUDY

You have been selected as a clinical interviewer (CI) for the Mental Health Surveillance Study (MHSS) of the National Survey on Drug Use and Health (NSDUH). The project staff welcomes you to the team for this important study and hopes you will find your responsibilities challenging, interesting, and enjoyable. We at Research Triangle Institute (RTI) look forward to working with you and appreciate the commitment and skill you bring to the project.

1.1 Research Triangle Institute (RTI)

Research Triangle Institute is an independent, nonprofit organization dedicated to conducting research that improves the human condition. With a worldwide staff of more than 2,800 people, RTI offers innovative research in survey and statistics, health and pharmaceuticals, advanced technology, education and training, and economic and social policy, international development, energy and the environment. Universities in North Carolina founded RTI in 1958 as the first scientific organization in and centerpiece of the Research Triangle Park, a science park located between the cities of Raleigh, Durham, and Chapel Hill, NC. In addition to its headquarters in North Carolina, RTI has eight regional offices within the United States and six offices around the world.

RTI's research is performed both in the United States and abroad under contract with federal, state, and local governments; public service agencies; and private-sector clients. In keeping with RTI's vision of becoming the world's leading independent research organization, the trade name has been changed to RTI International.[1] For survey work here in the United States, we continue to use both Research Triangle Institute and RTI because of positive respondent associations and name recognition.

1.2 Brief History of NSDUH

The National Survey on Drug Use and Health is currently an annual nationwide survey funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), an agency of the U.S. Public Health Service in the U.S. Department of Health and Human Services. These goals have been established for the NSDUH:

- to provide accurate data on the level and patterns of licit and illicit substance use;
- to track trends in the use of alcohol, tobacco products, and various types of other drugs;
- to assess the consequences of substance use and abuse;
- to identify groups at high risk for substance use and abuse;
- to monitor mental health problems, treatment and unmet need for treatment.

First conducted in 1971, the NSDUH has become the nation's leading source of information on substance use patterns and behaviors. Early on, the study was conducted at various intervals, settling into a pattern of about every two years. The demand for current, accurate information rose sharply by the early

[1] RTI International is a trade name of Research Triangle Institute.

1990s, prompting SAMHSA to conduct the survey annually starting in 1990. In 1992, the design shifted to a quarterly design where one fourth of the cases for the year are contacted and interviewed in each calendar quarter. Beginning in 1999, the sample size and design was expanded to allow for the reporting of drug use estimates for each of the 50 states and the District of Columbia. Beginning in 2002, the survey name was changed from the National Household Survey on Drug Abuse (NHSDA) to the National Survey on Drug Use and Health (NSDUH). The name was changed to more accurately reflect the survey's interest in the effects of drug use on users' mental health and overall health. Eliminating the term "abuse" from the title also projects a more positive, inclusive tone since researchers are interested in the experiences of drug users and non-users.

Research Triangle Institute has conducted the study since 1988. Over the years, RTI has revised or implemented new procedures designed to simplify and enhance the data collection process while maintaining the highest level of data quality.

The entire NSDUH data collection process is fully computerized. All screenings—which determine whom, if anyone, to interview in the household—are completed using a small handheld computer called a Hewlett-Packard iPAQ. Selected respondents are interviewed using Computer-Assisted Interviewing (CAI) on a Gateway laptop computer. Portions of the interview are conducted via computer-assisted personal interviewing (CAPI) where the field interviewer (FI) asks the questions and records the answers in the computer. The sensitive questions are completed using audio computer-assisted self-interviewing (ACASI), where the respondents read and/or listen to the questions and enter their own responses. With ACASI, even the interviewer does not know the responses to these personal questions. Studies repeatedly show that maximizing privacy helps encourage honest, accurate answers and produces high quality data.

Appendix A contains copies of materials and handouts often given to respondents by NSDUH field interviewers. Feel free to review these materials or visit the NSDUH website (http://nsduhweb.rti.org) to learn more about NSDUH.

1.3 Overview of the Mental Health Surveillance Study

The Mental Health Surveillance Study (also referred to as the Mental Health Study) was first conducted as part of the 2008 NSDUH. This study continues to be conducted in all states and is designed to assess the questions used within NSDUH's Mental Health module by:

- collecting data from all adult interview respondents through a Mental Health module within the NSDUH interview, and
- recruiting and completing in-depth follow-up telephone interviews with selected adult respondents.

For 2011, additional funding was provided to expand the MHSS from 500 follow-up telephone interviews conducted annually to 1,500 interviews. This expansion, referred to as the Expanded Mental Health Surveillance Study (EMHSS), will enhance researchers' ability to analyze and report on major mental health conditions (e.g., Post-Traumatic Stress Disorder, Manic Episode, Anorexia Nervosa, etc.) based on data collected from a nationally representative sample. While the MHSS and EMHSS have

separate funding sources, both studies follow the same protocols and procedures, and are conducted under the Mental Health Surveillance Study.

NSDUH field interviewers will be responsible for completing the initial NSDUH interview. For selected interview respondents, FIs will use special recruitment scripts displayed in the initial interview to recruit randomly selected respondents for a follow-up telephone interview.

As a clinical interviewer, you will be responsible for completing the follow-up telephone interview. Follow-up telephone interviews will be completed with selected adult respondents using a revised version of the Structured Clinical Interview for DSM-IV Axis I Disorders, also known as the SCID.

1.4 Study Sample

The Mental Health Study is being conducted in all 50 states and Washington, DC. Eligible persons will be the civilian, non-institutionalized population aged 18 years old and older. Persons aged 12 to 17 years old and respondents who complete the initial NSDUH interview in Spanish will not be eligible for the follow-up interview.

The MHSS sample will be embedded within the NSDUH sample. A sub-sample of English-speaking respondents aged 18 or older that complete the initial NSDUH interview will be selected to complete the follow-up telephone interview. The target number of completed follow-up telephone interviews for this year is approximately 1,500. Thus, each CI can expect to complete approximately 12-20 interviews per quarter.

1.5 Schedule

The Mental Health Study uses the same quarterly schedule as the NSDUH. Data collection will take place in all four quarters of the calendar year.

Given the cyclical nature of the schedule in that a new group of cases is released to field interviewers at the start of each calendar quarter, your workload will also be somewhat cyclical with more follow-up interviews generated and assigned at the beginning of the quarter than at the end. The graph on the following page illustrates the typical flow of total assigned follow-up interview cases through each quarter of the year.

Of course, it is best to complete the follow-up interview as soon as possible, preferably within the first week following the initial NSDUH interview. The vast majority of your assigned **follow-up interviews should be completed within 2 weeks** from the date of the initial interview. You have at most four weeks to complete the follow-up interview, so as needed, use the remaining 2 weeks for follow-up of difficult to contact respondents or persons who were out of town for a portion of the time period.

1.6 Organization

Because of the specialized nature of this study, the Mental Health Study organization structure has two types of supervisors for clinical interviewers. Knowing which type to contact for assistance is important.

You will report to one of two Data Collection Managers at RTI for all aspects of case assignment, management, completion, and reporting including case status and production, time, and expense information. Your Data Collection Manager will assign follow-up interview cases, monitor completion, and support you as needed to successfully complete your work.

For supervision of the follow-up interview task, that is, conducting the interview, asking the questions, documenting the answers and completing the SCID instrument, you will work with several Clinical Supervisors at RTI. These Clinical Supervisors will review your completed SCIDs, provide you with direct feedback, and answer any questions you may have about the proper administration of the SCID.

Please note that the Clinical Supervisors and Data Collection Managers work as a team to support your efforts while ensuring the collection of the highest quality data possible. ***Exhibit 1.1*** includes supervisory contact information. The chart also lists other contacts such as RTI project management, Headway (an RTI subcontractor) and RTI's Technical Support.

Exhibit 1.1 Project Staff Information

DATA COLLECTION MANAGERS			
Name	**Work Phone**	**Cell Phone**	**E-Mail**

CLINICAL SUPERVISORS			
Name	**Work Phone**	**Cell Phone**	**E-Mail**
Study Manager			

RTI MANAGEMENT			
Name	**Work Phone**	**Cell Phone**	**E-Mail**
Assistant Study Manager			
Operations Manager			

HEADWAY			
Headway Representative		FAX	

TECHNICAL SUPPORT GROUP (TSG) -

1.7 Project Abbreviations and Terminology

Throughout this handbook, and in other project materials, abbreviations are used. As a general aid for you, a list of these abbreviations is provided in ***Exhibit 1.2***. Some of the abbreviations and terms used have exact meanings or refer to specific project materials. These terms are briefly explained in ***Exhibit 1.3***, in alphabetical order.

1.8 Use of the Mental Health Study Handbook

This procedural handbook provides a detailed description of the tasks you will be required to complete for all aspects of the Mental Health Study, and is an excellent reference source. When searching for information, think about where the topic fits in the flow of work. Consult the detailed Table of Contents and the List of Exhibits. Using the key words shown there will help you narrow down your search.

Adherence to prescribed procedures and duties is extremely important to the success of the study. This handbook should be carefully studied before you attend training, as you prepare for work, and throughout data collection each quarter.

You may, however, have questions or encounter a situation for which you do not find an answer in this handbook. When in doubt about any situation, contact your Data Collection Manager.

Exhibit 1.2 List of Abbreviations

ACASI ..... Audio Computer-Assisted Self-Interviewing
CAI ..... Computer-Assisted Interviewing
CAPI ..... Computer-Assisted Personal Interviewing
Case ID ..... Case Identification Number
CBHSQ ..... Center for Behavioral Health Statistics and Quality (the SAMHSA office in charge of NSDUH and MHSS)
CI ..... Clinical Interviewer
CMS ..... Case Management System
CS ..... Clinical Supervisor
DCM ..... Data Collection Manager
DHHS ..... U.S. Department of Health and Human Services
DU ..... Dwelling Unit
ePTE ..... Electronic Production, Time, and Expense Report
EMHSS ..... Expanded Mental Health Surveillance Study
FI ..... Field Interviewer
FS ..... Field Supervisor
GQU ..... Group Quarters Unit
HU ..... Housing Unit
ID ..... Identification
INS ..... Immigration and Naturalization Services
MHSS ..... Mental Health Surveillance Study (Mental Health Study)
NHSDA ..... National Household Survey on Drug Abuse (past name)
NSDUH ..... National Survey on Drug Use and Health (current name)
PAPI ..... Paper and Pencil Interviewing
PT&E ..... Production, Time, and Expense Report
Q&A ..... Question and Answer Brochure
QuestID ..... Questionnaire ID
R ..... Respondent
ROC ..... Record of Calls
RTI ..... Research Triangle Institute
SAMHSA ..... Substance Abuse and Mental Health Services Administration
SCID ..... Structured Clinical Interview for DSM-IV Axis I Disorders
SDU ..... Sample Dwelling Unit
TSG ..... Technical Support Group
USPHS ..... United States Public Health Service

Exhibit 1.3 Definitions of Project Terminology

Audio Computer-Assisted Self-Interviewing (ACASI) – a type of computer-assisted interviewing used within the initial NSDUH interview to collect information from selected respondents for questions of a sensitive or personal nature. Respondents listen through headphones as the questions are read from computer audio files and enter the answers themselves directly into the computer.

Case Identification (Case ID) – a ten-character code that starts with the state abbreviation and uniquely identifies a dwelling unit selected for the NSDUH.

Case Management – a broad term used to describe the process of organizing, keeping track of and completing your work in a timely fashion.

CMS (Case Management System) – the web-based system that provides case assignment information and allows for the documentation of all work efforts through the Record of Calls *(see definition)*. Through the CMS, CIs also perform other study tasks such as documenting a distressed respondent situation.

Computer-Assisted Interviewing (CAI) – a generic term used to indicate that a computer is used during the NSDUH interview. It includes CAPI and ACASI *(see definitions)*.

Computer-Assisted Personal Interviewing (CAPI) – using a computer, the field interviewer reads the questions displayed on the computer screen to the respondent, then enters the response directly into the computer.

Consent – agreement to participate in a research study given by an adult or by a parent or guardian for his/her child. Giving consent indicates that he/she understands the meaning of, and has agreed to participate in, the study. The consent process used in any research project must be approved by a Human Subject's Committee or Institutional Review Board.

Electronic Production, Time, and Expense Report (ePTE) – CIs maintain an accurate record of daily production, time and expenses while working on this study and enter their information into the ePTE system. Payments for hours worked and reimbursement of expenses are based on the information submitted through the ePTE system. (In case of computer problems, paper PT&E forms are available as a backup).

Expanded Mental Health Surveillance Study (or EMHSS) – Additional funding provided to expand the MHSS from 500 follow-up telephone interviews conducted annually to 1,500 interviews for 2011. Both MHSS and EMHSS follow the same study protocols and procedures, and are fielded as the Mental Health Surveillance Study (see below).

Follow-up Interview – The follow-up telephone interview conducted with selected respondents who complete the initial interview. A clinical interviewer completes the follow-up interview, which asks questions about various mental health and other related issues.

Initial Interview – The NSDUH interview conducted with respondents, including respondents selected for the Mental Health Study follow-up interview.

Mental Health Surveillance Study (or Mental Health Study) – Nationwide study being conducted to assess the NSDUH's Mental Health module questions. This study is designed to collect data from all adult interview respondents through a Mental Health module within the NSDUH interview, and complete in-depth follow-up telephone interviews with selected adult respondents.

MH Case Manager – the system allowing CIs to easily audio record each interview through the laptop and then upload the audio file to a secure location at RTI.

Exhibit 1.3 Definitions of Project Terminology (continued)

Nonrespondent – a person who is eligible and selected for the NSDUH but who chooses not to participate.

Production Time and Expense (PT&E) – Backup paper forms used infrequently as needed to allow CIs to maintain and submit an accurate report of daily production, time and expenses while working on this study. Instructions for the proper completion of paper PT&Es are in ***Appendix B***.

Questionnaire ID (QuestID) – the identification code initially entered by the FI into the laptop in order to begin the initial NSDUH interview. The QuestID also represents that same respondent when assigned to a CI for a follow-up telephone interview.

Record of Calls (ROC) – a term referring to the system that allows CIs to record details about each and every attempt to contact a respondent. Details include the date and time, the status code, and any important comments.

Recruitment Scripts – After completing the initial NSDUH interview, field interviewers use a script programmed in the back-end CAPI portion of the interview to describe the follow-up interview to the respondent. The series of screens are referred to as "recruitment scripts" and it is the first time the follow-up interview is introduced to the respondent.

Respondent – a person who is eligible and selected for the study, and who agrees to participate.

Status Codes – two-digit codes used to indicate the current status of each case. These codes are recorded in the Record of Calls in the CMS, and are reviewed with your Data Collection Manager. When compiled into regular reports, these codes provide important information to project managers and the client on the progress of data collection.

Sample Dwelling Unit (SDU) – a dwelling unit [either a housing unit (such as a single family home or apartment) or a group quarters unit (such as a dormitory room or a bed in a homeless shelter)] that has been randomly chosen for inclusion in the NSDUH.

This page intentionally left blank.

2. YOUR ROLE ON THE MENTAL HEALTH STUDY

2.1 Introduction

Many factors make an interviewer successful. Some general guidelines are to follow **all** procedures and instructions carefully, know the study, listen and reply to respondents' needs and concerns, and maintain open communication with your supervisors.

This chapter outlines your responsibilities as a clinical interviewer (CI) for the Mental Health Study and shows in general how the above factors come together for a successful experience.

2.2 Your Responsibilities

For this study, NSDUH field interviewers (FIs) will complete the initial NSDUH interview with selected respondents. When prompted at the conclusion of the interview, the FI asks the respondent about participating in an additional follow-up telephone interview. For respondents who agree to participate, the FI obtains the respondent's first name, telephone number and suggested contact day(s) and time(s). The FI then pays the respondent $30 cash in advance, as a token of appreciation for participating in the follow-up interview.

As a CI, you are responsible for the next steps, including:

- receipt of follow-up interview case assignment;
- contacting the respondent; as needed, scheduling follow-up interview appointments;
- obtaining informed consent, including permission to audio record the interview;
- recording the interview (if permission granted);
- conducting the interview following all project procedures;
- evaluating possible respondent distress, handling the situation according to protocols;
- thanking the respondent for participating;
- copying the audio recording file to a secure website;
- shipping the completed and edited SCID to RTI; and
- reporting the status of each case promptly.

Exhibit 2.1 contains a diagram of the overall process.

Exhibit 2.1 CI Data Collection Process

Additional CI responsibilities include:

- read this handbook carefully prior to training;
- successfully complete the training program including assigned certification interviews;
- maintain the confidentiality of all survey data and materials at all times;
- maintain daily records of your data collection activities;
- review the status of cases during telephone conferences with your Data Collection Manager, providing detailed information about any problems with pending cases;
- review the technical aspects of SCID administration during calls with a Clinical Supervisor, accepting feedback to improve the quality of your interviews;
- complete and submit weekly electronic production, time and expense reports (ePTE reports) and other administrative forms in a timely manner and according to prescribed schedules;
- complete occasional training modules and/or exercises as assigned;
- meet or exceed project efficiency goals provided by the Data Collection Manager; and
- meet or exceed project quality standards (see ***Chapter 8***).

2.3 Professional Ethics and Respondents' Rights

Given your education and professional training, you are very familiar with professional ethics and respondents' rights. Therefore, the information in this section is not intended so much to educate you as to share RTI's specific policies related to these topics.

Ethics can be broadly defined as a set of moral values or principles of conduct governing an individual or a group. Organizations must show clients, employees, and the public, a prevailing sense of integrity, honesty, and responsibility in all aspects of work.

All survey research conducted by RTI is based on the highest ethical standards. Interviewers are expected to maintain the same professional ethics as all RTI researchers. RTI's professional reputation depends upon all employees and all interviewing staff making the commitment to ethical standards a high priority.

As part of professional ethics, the rights of survey respondents **must** be protected by all RTI and Headway personnel. These rights include:

- The <u>right of informed consent</u> refers to the legal requirement that respondents be given complete and accurate information so that they can make an informed decision about their participation in the survey.
- To ensure that all RTI studies meet the legal and ethical requirements of informed consent, all projects involving human subjects must be approved by our Office of Research Protection, which serves as RTI's Institutional Review Board (IRB) under federal regulations. This committee looks very closely at the written introduction to the study to be sure that the respondents are being properly informed.
- The <u>right to refuse</u> refers to an individual's right to decline to participate in the study or to refuse to answer individual questions once an interview has begun.

- The right of privacy is guaranteed by the federal **Privacy Act of 1974**. This Act prohibits the release of data gathered by or for a federal agency without the written consent of the respondent. Fines and penalties apply to individuals or organizations that violate this law. You can explain this to a respondent when trying to gain his or her trust.
- The right to accurate representation requires honesty in dealing with respondents and answering their questions about the survey. For example, you cannot tell the respondent that an interview will take only a few minutes if you know it will last about an hour, or say that you work for a government agency such as the Census Bureau or the Public Health Service.

All staff involved in the collection, processing, and analysis of the survey data must be continually aware of the important responsibility to safeguard the rights of the survey participants. As clinical interviewers, you are in direct contact with these respondents and must demonstrate high ethical standards in **all** project contacts.

2.4 Importance of Confidentiality

Any and all information you learn about respondents, whether directly from a response you receive or simply through casual communication during a telephone call must be treated as confidential. Because of the sensitive nature of the NSDUH subject matter, project staff members must take special precautions to protect the confidentiality of respondents.

- The confidentiality of all responses to the questions is protected under federal law by the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), a government-wide law that provides strong confidentiality protections to many federal agencies conducting statistical surveys. Under CIPSEA, all answers are used for statistical purposes only and cannot be used for any other purpose.
- The NSDUH is one of the few surveys where the full name of the respondent completing the interview is never recorded. While the respondents' addresses are known, this information is kept separate from the respondents' answers.
- Respondents should be reassured that any potentially identifying data, such as their address, are never made available to anyone outside the project staff.
- Individual responses are only analyzed in combination with other responses collected nationwide.

All NSDUH staff including project planners at SAMHSA, RTI managers and technical staff, and all interviewing staff must share the commitment to protect the confidentiality of the respondents. Under CIPSEA, the penalties for knowingly and willfully disclosing confidential information are a Class E Felony which includes imprisonment for up to five years and fines up to $250,000. Several NSDUH respondent materials outline the consequences of divulging respondent information to make it clear that their confidentiality will be protected.

Each year you must complete CIPSEA training, and then sign a Clinical Interviewer Confidentiality Pledge (shown in ***Exhibit 2.2***). By signing, you are entering into a contractual agreement that you will keep confidential all data you collect. If you have any questions regarding this policy, discuss them with your Data Collection Manager prior to signing.

Exhibit 2.2 Confidentiality Pledge

National Survey on Drug Use and Health (NSDUH)

CLINICAL INTERVIEWER CONFIDENTIALITY PLEDGE

Conducted by:
Research Triangle Institute (RTI)

Under contract to:
Substance Abuse and Mental Health Services Administration (SAMHSA)

Assurance of Confidentiality

RTI assures each respondent that the confidentiality of their responses to this survey's information request will be maintained by RTI and that no information obtained in the course of this activity will be used for any purpose other than the statistical purpose for which it was supplied. This assurance of confidentiality is required by the Privacy Act and supported by Section 511 of the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). The Confidential Information Protection and Statistical Efficiency Act of 2002 ensures the confidentiality of all information provided is protected under Federal Law and protects the privacy of research subjects by stipulating that all information collected shall be used exclusively for statistical purposes.

Under the authority vested in the Secretary of Health and Human Services by Title V of the Confidential Information Protection and Statistical Efficiency Act of 2002 (PL107-347), all persons who --

1. are employed by RTI and its contractors and cooperating agencies; and
2. have, in the course of that employment, access to information which would identify individuals who are the subjects of a research project referred to as "National Survey on Drug Use and Health,"

are required to protect the privacy of the individuals who are the subjects of that research by withholding their names and other identifying characteristics from all persons not connected with the conduct of that research.

Agreement and Pledge

I have carefully read and I understand the Assurance of Confidentiality that pertains to the confidential nature of all data to be handled in regard to this project. As someone who will have access to data from the NSDUH, I pledge that I will not disclose any confidential information obtained under the terms of this contract to anyone other than authorized project staff at RTI or SAMHSA. I understand that under the Confidential Information Protection and Statistical Efficiency Act of 2002 [Section 513] I am subject to criminal felony penalties of imprisonment for not more than five years, fines of not more than $250,000, or both, for violation of the Act. I have also completed and fully understand the CIPSEA training provided to me.

I further understand and agree to comply with the following confidentiality provisions:

1. Any materials that would permit the identification of survey participants are to be treated as confidential. These include both hardcopy and electronic records.
2. When confidential records are in use, access to all such records is limited to persons who have signed the project Confidentiality Pledge.
3. Confidential records must be kept in a secured (locked or password-protected) location when not in use.
4. Information obtained from the data collected or used under this contract may not be released to unauthorized persons.
5. Any involvement in the study is for research purposes only. Interviewers cannot and will not utilize their clinical knowledge or skills to provide medical or psychiatric advice, diagnosis or treatment to respondents.
6. No data, tabulations, or analyses obtained under this contract may be released or used without prior written approval of SAMHSA. At the close of the contract all data must be returned to SAMHSA.
7. Any breach of confidentiality must be reported immediately to the RTI Project Director. This information will be shared with the SAMSHA Project Officer.
8. Obligations under this agreement will survive the termination of any assignment with RTI.

_____	_____
Name (Print)	*RTI ID #*
_____	_____
Signature	*Date*

Disposition: Original to the NSDUH Project File maintained by the Project Secretary.
Revised: November 21, 2007

2.5 Protecting Both Confidentiality and Study Integrity

In order to protect the confidentiality of respondents and preserve the scientific integrity of the Mental Health Study, there are several important protocols to follow regarding your interactions with respondents:

- Avoid calling the respondent by name (once you have confirmed you are speaking with the correct person).
- Never write down any potentially identifying information on the paper SCID while you are conducting the interview.
- Remind respondents, if needed, to not say their own name or potentially identifying information such as a place of business or address during the interview.
- Do not identify or introduce yourself as Dr. Jane Smith, but simply as Jane Smith.
- Do not offer information about your training or specialized background. If a respondent asks whether you are a doctor, psychiatrist, psychologist, social worker, or other mental health professional, you may disclose that you have medical or psychological training that allows you to fully understand the questions you are asking. Explain that your involvement in the study is for research purposes only.
- **Do not** provide psychological or medical counseling, advice, or other mental health services, even if you are qualified to do so. If asked for advice/diagnosis/treatment, refer the respondent to the national Lifeline.

Additionally, always consider the security of the confidential data you are collecting.

- Be sure the paper SCID is stored in a securely locked place when not in use, such as in a locked file cabinet.
- Since the laptop contains audio recordings of interviews, press the Ctrl/Alt/Del keys simultaneously to lock the computer or turn off the laptop when you are not working so that others cannot use it (see ***Section 3.2.3***).
- Log out of the web-based case management system when finished, to maintain security.

2.6 Materials, Supplies, and Equipment

There are a variety of materials, supplies, and equipment you will use to conduct your work. Adequate quantities of materials and supplies will be sent to you prior to your data collection activities. You must use the proper materials for each follow-up interview. Be sure to monitor your level of supplies for upcoming work and if you require additional supplies, contact your Data Collection Manager. Descriptions of the purpose and use of each item can be found throughout this handbook.

Materials Needed for Interviewing:

- NSDUH laptop computer equipment, including power cords and headphones
- One of two recording controls (telephone handset or wall jack) to use when audio recording the interview
- SCID booklet
- Pens to use with the SCID, either blue or black ink

- Mental Health Study CI Handbook
- SAMHSA Calling Card

Administrative and Other Materials

- Pre-printed Confidentiality envelopes for SCID
- Prepared FedEx labels to RTI and to RTI Technical Support
- FedEx packages/envelopes
- Working copy ePTEs
- Headway Expense Report Form
- Paper PT&Es (for backup)
- Pens, pencils,

This list of equipment and materials includes everything you need to conduct interviews according to project protocol. If you feel the need for some additional item or handout, please forward your suggestion to your Data Collection Manager who can pass it along to the appropriate staff for consideration. Please do not create your own materials.

At the conclusion of data collection, make arrangements to properly dispose of any remaining study materials as instructed by your Data Collection Manager, by either shipping them back to RTI or by recycling or trashing the documents.

Also, when finished with your Mental Health Study assignment or as directed by your Data Collection Manager, return the laptop and its accessories to RTI Technical Support. Use the packaging provided to safely pack the equipment, and then ship it via FedEx using the supplied, preaddressed label to Technical Support. ***Section 7.8*** contains detailed FedEx instructions.

This page intentionally left blank.

3. EQUIPMENT

3.1 Introduction

Because all follow-up interviews for the Mental Health Study will be conducted over the telephone, RTI will provide you with a calling card to use for contacting respondents and conducting interviews. RTI also provides a Gateway laptop to use for e-mail and connecting to the Internet for assignment and reporting information. In addition, all telephone interviews will be recorded (with the respondent's permission) using the laptop connected to your telephone via one of two recording controls. These audio recordings will be reviewed along with the SCIDs to assure the data collected are of high quality. **Remember, respondents must give their permission before the interview can be recorded**. Information regarding the procedures for obtaining respondents' permission to record the interview is in ***Chapter 4***.

3.2 The NSDUH Laptop

You will use a NSDUH laptop computer to perform a number of your duties including accessing the Case Management System (CMS), checking e-mail, and audio recording interviews through the MH Case Manager. The next sections review the initial laptop set up and the login process. As needed for reference, consult ***Appendix C*** for details regarding the laptop hardware.

3.2.1 Initial Set Up

During training, you will learn and practice each of the following steps in the initial set up of your laptop. Refer to Appendix C for additional details.

1. Connect the laptop to electrical power (see ***Section C.4.1***).

2. As needed, connect the laptop to your high speed internet connection. If using a wireless connection, be sure to use a secure network.

3. Plug the headphones into the green headphone jack on the left side to disable the laptop speakers.

4. Connect the laptop to the telephone.

<u>For an ANALOG PHONE LINE</u>, use the Wall Jack Recording Control (pictured below) which goes between the wall and the telephone and follow these steps:

4a. Be sure the switch on the recording control is set to "Record" (not "Play").

4b. Disconnect the phone cord from the wall jack and plug into the control.

4c. Connect the recording control to the wall jack using the phone line attached to the controller.

4d. Connect the recording control's audio plug to the Mic jack on the front of the laptop, which is the jack to the left of the headphone jack near the right hand speaker.

For a DIGITAL PHONE LINE, use the Handset Recording Control (pictured to the right) which goes between the telephone and the telephone handset. Please note that this control does NOT work with a cordless telephone. Recording takes place through the handset of the telephone; therefore you must always speak through the telephone handset.

To connect, follow these steps:

4e. Be sure the switch on the recording control is set to "Record" (not "Play").

4f. Disconnect the coiled cord from the handset jack on the telephone and plug it into the jack labeled "Handset" on the recording control.

4g. Plug the short coiled cord from the recording control into the handset jack on the telephone.

4h. Connect the recording control's audio plug into the Mic jack on the front of the laptop, which is the jack to the left of the headphone jack near the right hand speaker.

Note: Due to electrical interference between the Handset Recording Control and the AC Adapter for the laptop, you may notice a buzzing sound. This sound can be minimized if you disconnect the laptop from electrical power and run the laptop solely on battery power. To prevent loss of audio data, BE SURE you have adequate battery power to keep the laptop running for the entire interview. See ***Section C.4*** for more information on battery power and low battery warnings.

3.2.2 Login Procedures

Once your laptop computer is correctly connected, begin the login process to establish the proper connection with RTI.

1. Press the power button to turn on the laptop. Wait patiently as it powers on.

2. The PointSec security password prompt appears:

3. Carefully enter your user name (supplied separately) and the case-sensitive password. Note: The Touchpad does not work on this entry screen, so press TAB to move between the fields. When finished, press Enter.

4. The next pop-up box indicates your log in was successful. Press Enter.

5. At the Windows Login screen, use the same user name and password. The user name is prefilled, so type your password and press Enter.

6. You will next arrive at the Desktop. Icons on the desktop include:

   My Documents
   My Computer
   Recycle Bin
   Internet Explorer
   CMS Shortcut (see ***Section 3.3***)
   E-mail Shortcut (see ***Section 3.4***)
   Sound Volume
   MH Case Manager (see ***Section 3.6***)
   ePTE Shortcut (See ***Section 7.5***)

3.2.3 Locking the Laptop

Because your laptop contains confidential information, access MUST be limited to only project personnel. Each time you step away from the laptop, you must lock it to protect the information it contains.

1. To lock the laptop, simultaneously press the Ctrl/Alt/Delete keys.

2. At the Windows Security screen, click Lock Computer. As needed, click Cancel. Note: Do NOT use the Change Password option from this screen.

When ready to resume work, simply press Ctrl/Alt/Delete again. Enter your Security Login password as prompted.

3.2.4 Shutting Down the Laptop

When you have finished working for the day, go ahead and shut down the laptop. Doing so increases security and saves energy as well as wear on the laptop.

1. To shut down the laptop, simultaneously press the Ctrl/Alt/Delete keys, and then choose Shut Down.

OR

2. Click the Start button, select Shut Down, and then click OK.

3.3 Accessing the CMS

The CMS is the central point for many of your Mental Health Study tasks including case assignment and reporting.

1. Be sure you are connected to the Internet either via an Ethernet cable or a wireless connection. To check, open Internet Explorer and see if you can view a standard web page such as yahoo.com.

2. Double click the CMS icon on the desktop. The website address is:

3. Enter the CMS user name and password (provided separately).

4. Then enter your individual CI user name and password to access the site.

5. The CMS Main Page appears. On the CMS Main Page are several different options allowing you to perform various tasks:

Link	Tasks	Reference
Case Status	Check assignment Update case status	Section 4.2.2 Section 6.2.4
Email	Link to RTI e-mail	Section 3.4
Incident Report Form	Document any distressed respondent situation	Section 5.7.4
Change Password	Update password, as needed	For use as needed Consult Data Collection Manager

3.4 RTI E-Mail

To ensure the secure exchange of information through e-mail, you will receive an RTI e-mail address and password. To access the RTI e-mail system, follow these steps:

1. With the laptop connected to the Internet, use the shortcut on the Desktop, or go to the following link: You can also access the system through the link on the Main Page of the CMS.

2. Next, you will see the screen shown below. Enter your full *Email Address* and *Password* in the fields provided, and click **Sign In**.

3. The next screen you see will be your Inbox.

The RTI e-mail system contains many typical e-mail features. Working knowledge of typical e-mail functions is required. Should you have any questions or problems using the RTI e-mail system, feel free to contact your Data Collection Manager.

3.5 Calling Card

RTI will provide you with a calling card for your work on the Mental Health Study. This card should be used for all calls you make for the Mental Health Study that are not toll-free calls. Follow these steps to use the calling card:

5. Dial *67 on your telephone. (*Optional*: This step blocks your phone number from displaying on Caller ID)
6. Dial the toll-free number on the back of the card
7. Enter the card number (printed on the front and back of the card)
8. Enter 0 + Area Code + Telephone number you wish to call

Please remember Step 1 which is not included with the instructions for placing a call printed on the back of the calling card. This optional step blocks the phone number from which you are calling from being displayed on a Caller ID feature that may be on the respondent's telephone. If you are unable to successfully place a call to a respondent because of a privacy blocker function on his/her phone system and you do not feel comfortable unblocking your number, notify your Data Collection Manager.

3.6 Audio Files

As an important quality control measure, you will ask permission and if granted, will make an audio recording of each interview. By connecting your telephone through the laptop, special software on the laptop then records the conversation. When finished, you can upload this audio file to a secure RTI website accessible only by qualified RTI personnel.

The MH Case Manager available on the laptop controls all aspects of the audio recording process. The following sections provide details on using the MH Case Manager.

3.6.1 Connecting and Using the MH Case Manager

Follow the steps below to use the laptop to record audio files:

1. Be sure the laptop is properly connected to the telephone. Refer to ***Section 3.2.1*** as needed.
2. Be sure you are connected to the Internet either via an Ethernet cable or a wireless connection. To check, open Internet Explorer and see if you can view a standard web page such as yahoo.com.

   Note: If you are unable to connect to the Internet, you can still record an interview. See the end of ***Section 3.6.3*** for directions.
3. Double click the Case Manager icon on the desktop.
4. Enter the same main CMS user name and password used to access the CMS.
5. The Case Manager screen appears, listing available cases and various task buttons.

The chart shows the identification number for each case—the QuestID—along with other respondent contact information. For details, see ***Section 4.2.2***.

6. As the number of files increases and eventually fills the screen, you may have to scroll through the list to find a specific case. By default, cases are displayed in numerical order by QuestID. To change the order, click the column heading you wish to sort by. For example, if you wish to have the more recent files showing, click Create Date.

3.6.2 Initial Test Audio Files

Before contacting a respondent, you must first perform a quality control check to ensure the laptop is properly recording audio files. Each time before calling a respondent, follow the steps below to complete an initial test audio file:

1. Highlight the appropriate QuestID for the case. The information for the selected case displays towards the bottom of the screen.

2. Click "Recording Test" in the lower left corner. The following screen appears:

3. Click the green "Start Recording" button.

4. Make your test call by calling a local number with an automated voice. Listen for that voice, and also say a few words yourself. One suggestion is to call a work or cell phone which will be answered by voice mail.

5. When finished, click the red "Stop Recording" button.

6. On the next screen, click Play to listen to the recording. Be sure you can hear both sides of the conversation. Since the headphones connected to the laptop disable the laptop speakers, use the headphones to listen to the recording.

7. When the audio test file finishes playing, answer the question "Did the recorded audio correctly capture the conversation?" Click Yes or No as appropriate. For a No response, you must complete another test recording before contacting the respondent.

8. When finished, click "Close."

3.6.3 Recording the Interview

After properly completing the test recording, you can contact the respondent (see ***Chapter 4***) and with permission, record the interview.

1. The "Start Recording" option becomes active following the successful test recording. So once you have permission to record, click the green "Start Recording" button.
2. The button turns red and says "Stop Recording" as shown below. You will see blue flashes along the bottom of the screen, indicating that recording is taking place. All the fields fill automatically.

Also, the audio file name appears in the bottom left corner of the screen under the "File Name" column. The audio file is automatically named, and includes your interviewer ID, the QuestID, as well as the date and time the file was recorded.

3. When finished with the interview, hang up the telephone and click the red "Stop Recording" button.

   Note: Should you need to stop the interview and start again, please note that you will have two audio files associated with the case. Be sure to upload both files to RTI.

It is possible to record an interview when you are **unable to connect to the Internet**. While this is not preferred, it is possible. Follow these steps:

1. When accessing the MH Case Manager, click "Cancel" in the user name and password pop-up box.
2. Enter the QuestID for the interview in the "QuestID" box at the bottom of the screen.
3. Enter your interviewer ID when prompted.
4. Complete the initial test recording and then record the actual interview.
5. Please note you must be properly connected to the Internet to upload the audio file to RTI.

3.6.4 Uploading Audio Files to RTI

After completing each interview (and before starting another interview), upload the audio recording to RTI through the MH Case Manager, so that the files can be accessed by authorized project staff. After hanging up with the respondent and clicking "Stop Recording," follow these steps to upload the audio files:

1. In the second to last column on the right, notice the response in the "Uploaded?" column is blank, meaning the file has not been sent to RTI but has been saved on the laptop.

2. Click the "Upload" button located to the right of the green "Start Recording" button. All interviews that have not yet been sent to RTI will be uploaded.

3. When finished, the "Uploaded?" column has a check indicating that the file has been uploaded to RTI. The Upload Date/Time column is filled as well.

NOTE: You must be connected to the Internet to upload successfully. If you get an error message, first check your Internet connection (minimize the Case Manager and click on Internet Explorer to see if you can view a web page such as yahoo.com). If you are connected, return to the Case Manager and try to upload again.

The audio files will remain on the laptop's hard drive. You will not have to delete these files.

3.7 Technical Support Contact Information

If you have any problems using the audio recording process or uploading files, first contact your Data Collection Manager. If he/she is unavailable or directs you to do so, contact the NSDUH Technical Support Group (TSG) at . Technical Support is available Monday through Friday from 8:30 am to midnight (Eastern time) and from 10:00 am to midnight (EST) on weekends.

Explain that you are a clinical interviewer working on the **Mental Health Study** and you have a problem requiring their assistance.

4. CONTACTING RESPONDENTS

4.1 Introduction

With every scientific survey, many components contribute to the overall success of the research. As a clinical interviewer for this study, you are responsible for one of the most important aspects of the study: making sure the follow-up interview is administered properly. The interview process begins with preparation tasks as explained in this chapter including receiving cases, initially contacting respondents, and completing the study introduction process according to protocol. ***Chapter 5*** discusses the steps required while conducting an interview and offers suggestions for dealing with some respondent situations you may encounter.

4.2 Case Assignment

This section describes the process of learning details about a case assigned to you for completion.

4.2.1 Case Identification Information

Each household selected for an initial NSDUH interview is assigned a Case Identification number (Case ID) which is a 10-character code that starts with the state abbreviation and uniquely identifies that household for the study (e.g. NY10010001). When the initial interview is generated and completed, a 7-digit number called the Questionnaire ID (Quest ID) is assigned to that particular interview respondent (e.g. 9994435).

Because you will be conducting follow-up interviews with NSDUH respondents who have already completed an initial interview, the cases you are assigned will be identified by the 7-digit **Quest ID**. All cases will be stored and tracked on a secure RTI web site that will be accessible to you and NSDUH project staff. All clinical interviews must be completed within four weeks from the time of the initial interview with the vast majority of **cases completed within two weeks**, leaving the final two weeks for continued follow-up of difficult-to-contact cases. All clinical interviewing work must be completed by the deadlines provided. The Data Collection Manager will assign cases based on the time the initial interviews are completed and your availability.

4.2.2 Case Assignment Details

On each day you have indicated you are available to work, check for new cases assigned to you. The web-based Case Management System (CMS) provides important details about new cases including contact information for the respondent with whom you will be conducting the follow-up interview.

To check for new cases, first log into the CMS website (refer to ***Section 3.3***, Accessing the CMS, for details).

To see assignment information, click the "**Case Status**" link on the CMS Main Page.

At the top of the screen, a table allows you to filter the information displayed.

To easily view your cases, simply click "Filter Data."

A listing of cases assigned to you and additional information associated with each case appears.

10 Records shown below.

Case Data Information for CS (902)

QuestID	Status Code	Qtr	Created Date	Contact Name	Contact Phone	Contact Note	Time Zone	State	Best Days	Best Time	DateTime Note	CI	Case Note	Address
9999991	00 10/03/2010	4	10/03/2010	Inez	(999)-999-9990	(999)-999-0001 work	Eastern	NC	M, W	7-9 pm		SMITH JANE (676882)	No	Click here
9999992	00 10/04/2010	4	10/04/2010	Keith	(999)-999-9991		Eastern	IN	Tuesday	afternoon	weekends also	SMITH JANE (676882)	No	Click here
9999993	00 10/05/2010	4	10/05/2010	Jasmine	(999)-999-9992	(999)-999-0002 cell	Mountain	CO	W or Th	6 or 7 pm	works late nights, sleeps days	SMITH JANE (676882)	No	Click here
9999994	00 10/06/2010	4	10/06/2010	Rosa	(999)-999-9993	Cell phone only	Pacific	CA	weekends	any time	weekends when cell minutes free	SMITH JANE (676882)	No	Click here

The information displayed in each column is defined below.

Quest ID: A unique identifying number used to identify the case

Status Code: A code that indicates the current status of the case (refer to ***Section 6.2***) along with the entry date for the code

Qtr: Indicates the Quarter in which the case is generated

Created Date: Displays the date of the initial interview

Contact Name: Interview respondent's first name

Contact Phone: Interview respondent's phone number

Contact Note: Notes made by the field interviewer about the respondent's name/number including an alternate contact number

Time Zone: Time Zone for the respondent's area

State: Respondent's home state

Best Days: Best day(s) to reach the respondent, as reported by the respondent during the initial interview

Best Time: Best times to reach the respondent, as reported by the respondent

DateTime Note: Notes made by the field interviewer about the contact day/time information

CI: Assigned Clinical Interviewer name and ID number for the case

Case Note: "Yes" indicates notes about the case have been entered by the Clinical Interviewer and/or the Data Collection Manager (click "Yes" to display notes)

Address: Link to respondent's address. This should only be used in the case of a distressed respondent and according to the Distressed Respondent protocol (see ***Section 5.7.4***).

The columns most useful to you when obtaining your assignment information are the **Quest ID** column, and the **Contact Name, Contact Phone, Best Days** and **Time** columns which display the respondent's first name, phone number and the best day and time to contact. Any notes provided by the field interviewer may also be informative. The **Time Zone** column lets you know the time of day in the respondent's area. Be sure to also note the **Created Date** column as interviews should be completed within 2 weeks, and must be completed within 4 weeks of this date.

Due to these deadlines as well as the importance of contacting the respondent soon after the completion of the initial NSDUH interview, do not wait more than a day or so to contact the respondent. For example, if the best day/time listed for a newly assigned case is 5 days away, go ahead and contact the respondent. Perhaps you can establish a firm appointment to complete the interview at that best day/time, or maybe you can go ahead and complete the interview when you call.

4.3 Contacting the Selected Respondent

The interview respondents you will be contacting are familiar with the purpose of your call as they will have already agreed to participate in this special study. However, they may have questions or need additional information. Strategies and tools for contacting selected respondents are outlined here.

4.3.1 Preparation

Before you dial the number, review the list of interviewing materials in ***Section 2.6*** to make sure you have all necessary paperwork and equipment. Also be sure your laptop is connected to the telephone, ready to record the audio, and that you have completed the initial test recording through the MH Case Manager (see ***Section 3.6*** for details).

Keep in mind the subtle differences when interviewing over the telephone.

Location: To ensure privacy for your respondent, be sure you are in a private location within your home or office. Even though you are not to call the respondent by name during the interview, other identifying information may be mentioned, such as asking for the respondent or verifying the telephone number when you call.

Speakerphone: Do NOT use a speakerphone while conducting the interview, as that greatly increases the chance the respondent's answers could be overheard. To increase your comfort level while interviewing, you may use a headset with the Wall Jack Recording Control used with analog phones (refer to ***Section 3.2.1***).

Style: Think about your telephone style. You should be serious, pleasant, and self-confident. Your voice and words must convey your credibility; it is not just ***what you say*** but ***how you say*** it. For example, if you sound uncertain, this feeling will be communicated to the respondent, who will react accordingly.

Voice: Be sure to speak clearly, and at a comfortable, conversational pace. Based on RTI's experience, the first twenty seconds of your telephone call can determine the outcome, so be prepared.

Communicate: Remember to ***talk to*** the respondents, ***not at*** them. This requires that you respond interactively and listen to what the respondents say. If they believe you are really interested in them, they will be more likely to participate fully.

Always keep in mind that you are a professional doing a very important job. As a professional, you are expected to possess a great deal of knowledge about the survey you are conducting—its purpose, the type of sample, the interview process, etc. Also, you must always maintain the highest of ethical standards, collecting data with complete objectivity and treating with the utmost confidentiality and respect **all** information gathered or observed during an interview. You must convey to every respondent that you are a professional, that you are completing the interview in a completely confidential manner, and that you are not affected by any personal biases, opinions or prejudices.

4.3.2 Initial Contacts

Often your first call establishes a connection between you and the respondent as well as an appointment time for the interview. To maximize efficiency and save time for both you and the respondent, be prepared and attempt to complete the interview during your first contact.

An introductory script contained within the first few pages of the SCID booklet (***Exhibit 4.1***) guides you through the introduction process as well as the necessary steps before beginning the interview (refer to ***Section 4.4***).

Exhibit 4.1 Clinical Interviewer Introduction Script

Before you call, **be prepared**:

- Review the assignment information provided including the respondent name, telephone number, as well as the date of the initial interview.
- Have your schedule available (in case you need to schedule an appointment).
- Have all interviewing materials available.

VERIFY NUMBER AND LOCATE RESPONDENT

Hi, my name is _____ and I'm calling on behalf of the U.S. Public Health Service. Is this [PHONE NUMBER]?

YES: PROCEED BELOW.
NO: I apologize. I need to double check my records. Thank you for your time. END CALL.

I'm trying to reach [FIRST NAME] who agreed to take part in a telephone interview we're conducting. May I speak to [FIRST NAME]?

IF R NOT HOME OR UNAVAILABLE

When would be a good time to call again? ENTER CODE 51 AND DETAILS IN CMS.
Thank you for your time. END CALL.

IF R AVAILABLE

(Hi, my name is _____.)
You recently completed an interview in your home with an interviewer working on the National Survey on Drug Use and Health. I am the interviewer you were told would contact you for a follow-up telephone interview. Do you recall completing the first interview?

YES: PROCEED BELOW.
NO: VERIFY FIRST NAME OF PERSON YOU ARE SPEAKING TO.
IF NOT SPEAKING TO CORRECT RESPONDENT, ASK TO SPEAK TO RESPONDENT.

IF NAME IS CORRECT AND RESPONDENT DOESN'T RECALL INITIAL INTERVIEW, REMIND OF DATE OF INITIAL INTERVIEW.

IF CORRECT RESPONDENT STILL NOT FOUND: I apologize. I need to double check my records. Thank you for your time. END CALL. ENTER CODE 59 AND INVESTIGATE.

Are you in a place where you can safely talk on the phone and answer my questions?

YES: PROCEED
NO: Are you able to move to a place where you can safely talk?
YES: PAUSE, THEN CONTINUE
NO: When would be a good time to call again? ENTER CODE 50 AND DETAILS IN CMS. Thank you for your time. END CALL.

Is now a good time to complete this interview?

YES: PROCEED. BE SURE TO READ VERBATIM.
NO: When would be a good time to call again? ENTER CODE 50 AND DETAILS IN CMS.
Thank you for your time. END CALL.

Exhibit 4.1 Clinical Interviewer Introduction Contact Script (Continued)

PRIVACY

Because you may not want others to hear the responses to some of our questions, I'd like to be sure you're in a private area. Where are you right now? Are you at home, at work, or somewhere else? Are you in an area where you can answer these questions privately?

YES: PROCEED
NO: Please consider moving to a more private area. Do you need more time?
YES: PAUSE, THEN CONTINUE
NO: CONTINUE

INFORMED CONSENT

Before we begin, I would like to remind you of the study details. This study, sponsored by the United States Public Health Service, asks questions about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Although there is no benefit to you personally, knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States. While the interview has some personal questions, federal law keeps your answers private. The only exception to this promise of confidentiality is if you tell me that you intend to seriously harm yourself or someone else; in this situation I may need to notify a mental health professional or other authorities.

Your participation is voluntary. You may consider some of the questions to be sensitive in nature and some of the questions may also make you feel certain emotions, such as sadness. Remember that you can refuse to answer any questions that you do not want to answer, and you can stop the interview at any time. If you become upset at any time during the interview and wish to speak to a mental health professional about how you are feeling, I will provide you with the toll-free hotline numbers that are printed on your payment receipt from the first interview. It is important for you to keep in mind that I will not be providing you with a psychological diagnosis or any mental health advice or counseling. The information we are collecting today is only for research purposes.

These study details are also included on the Follow-up Study Description you received from the interviewer who met with you in your home. Do you have any questions before we begin? ANSWER ANY RESPONDENT QUESTIONS.

Is it OK to continue with the interview?

YES: PROCEED TO NEXT PAGE
NO: BASED ON CONVERSATION:
What sort of concerns do you have about participating?
OR
Are there other questions that I could answer for you?

IF R STILL UNWILLING TO PARTICIPATE: Thank you for your time. END CALL.

Exhibit 4.1 Clinical Interviewer Introduction Contact Script (Continued)

RECORDING PERMISSION

In order to ensure that I am conducting this interview accurately and properly, I would like to make an electronic audio recording of this interview. This is done strictly for quality control purposes. The recording will only be listened to by staff members on the project who have signed confidentiality pledges. The recording will be stored in a secure manner and will not contain your name—only a random number that will be assigned to this case. To help maintain confidentiality, we ask that you not give your name or any other identifying information, such as an address or place of business, during the interview. All recordings will be permanently destroyed within eighteen months after the end of the data collection period. You can still do the interview if you do not want me to record it.

Do you agree to allow me to record the interview?

YES: I will now begin recording. START RECORDING AND SAY: "This is [YOUR FIRST AND LAST NAME] conducting telephone interview [QUEST ID] on [DATE]."
NO: DON'T RECORD

Ok, let's get started.

CI NOTES:
IF ASKED AT ANY TIME BY A RESPONDENT WHETHER THE INTERVIEWER IS A DOCTOR, PSYCHIATRIST, PSYCHOLOGIST, SOCIAL WORKER, OR OTHER MENTAL HEALTH PROFESSIONAL, YOU MAY DISCLOSE THAT YOU HAVE MEDICAL OR PSYCHOLOGICAL TRAINING THAT ALLOWS YOU TO FULLY UNDERSTAND THE SURVEY.

HOWEVER, YOU SHOULD EXPLAIN THAT YOUR INVOLVEMENT IN THIS STUDY IS FOR RESEARCH PURPOSES ONLY AND IN NO WAY CONSTITUTES MEDICAL OR PSYCHOLOGICAL ADVICE, TREATMENT, OR DIAGNOSIS. EXPLAIN THAT THIS IS NOT THE NATURE OF THIS EFFORT.

IF RESPONDENT REQUESTS PSYCHOLOGICAL COUNSELING OR ADVICE OF ANY KIND, REFER HIM/HER TO THE NATIONAL LIFELINE. IF RESPONDENT IS INTERESTED IN CONTACTING THE LIFELINE, OFFER TO STAY ON THE PHONE AND CONNECT THEM VIA A THREE-WAY CALL.

4.3.3 Follow-Up with Respondents

If the respondent cannot complete the interview during your initial call, try to set a firm appointment with him/her for the interview. At a minimum, ask for a good time of day to call back. Document all details of the contact in the CMS, including appointment information (see ***Section 6.2***). Review all contact information prior to calling again.

If you get an answering machine or voice mail, listen carefully to the message as it can help you verify the name and/or telephone number. On your first call, leave a message such as:

> "Hi. I'm [YOUR NAME] and I'm calling on behalf of the U.S. Public Health Service, trying to reach [RESPONDENT'S NAME] about an important study. I am sorry I missed you. I will call again. Thank you."

If you continue to call and only get voicemail, leave messages every third call and for any call made during the time specified by the respondent as a good time to call.

In general, it is best to not leave a call back number for security reasons. However, if you are having trouble reaching a respondent, you may consider leaving a call back number. For privacy reasons, we recommend not using your home number. Also, you should not contact a respondent using a cell phone, since you would be unable to conduct/record the interview. If you leave a work number on the voice mail, the respondent might call you at an inconvenient time for you to conduct the interview, although you could perhaps establish an appointment. Consider providing that number only for difficult cases and be sure to discuss the situation with your Data Collection Manager.

4.4 Beginning the Clinical Interview

Once you have the respondent on the phone ready to begin the interview, you have a few more steps to complete before you can begin the actual interview. First, as scripted on the introductory script (refer to ***Exhibit 4.1***), remind the respondent about the importance of privacy.

4.4.1 Informed Consent

Even after a respondent has initially agreed to participate, you must follow all informed consent procedures. The respondents' Right to Informed Consent is a critical part of any legitimate research project, as each person must receive all the information necessary to make a completely informed and knowledgeable decision about participating. Even if the respondent immediately agrees to be interviewed, you are still required to follow the informed consent procedures and read the scripted introduction before you begin the interview. If a respondent interrupts while you are reading the informed consent script, thank him/her for his/her willingness to participate, but explain that you are required to read certain information to make sure he/she is fully informed prior to agreeing to participate. This information **must** be made available to each respondent. By following the scripted introduction (refer to ***Exhibit 4.1*** as needed), you will properly cover all informed consent requirements.

4.4.2 Explaining the Survey and Answering Questions

Although the respondent has already agreed to participate and received a Follow-up Study Description (see ***Exhibit 4.2***) and the additional cash payment for the telephone interview, you may still encounter respondents with questions or concerns about participating in an additional interview.

Be sure you are familiar with the background information on the study (see ***Chapter 1*** and ***Appendix A***). We encourage you to learn as much as possible about the NSDUH so that you are comfortable answering questions about it. Specific questions relating to the Mental Health Study interview and suggested responses are given in ***Exhibit 4.3***.

4.4.3 Recording the Interview

As a quality control measure, you are to create an audio recording of each interview. (See ***Chapter 3*** for details on the audio recording process.) During the introduction, you must review the recording process with each respondent and ask for permission to record. If the respondent agrees, start the recording, read the statement found on the introductory script identifying you as the CI and the Quest ID for the case, then begin the interview.

If the respondent does not give permission for recording, do not start recording but proceed with the interview. When later entering the status for the case, include in the Notes the respondent's reason for not giving permission to record, if you know the reason.

Exhibit 4.2 Follow-Up Study Description

U.S. Public Health Service
Center for Behavioral Health
Statistics and Quality

Follow-up Study Description

You have been randomly chosen for this special study for the 2011 National Survey on Drug Use and Health. This study, sponsored by the United States Public Health Service, will ask questions about various mental health issues such as depression, anxiety, post traumatic stress disorder and substance dependence. Although there is no benefit to you personally, knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.

If you agree to participate in this follow-up interview, your first name and telephone number will be collected but will be used only for re-contact purposes. Your name and telephone number will not be included on the interview forms on which your answers will be written, or on any interview audio files that might be recorded. While the interview has some personal questions, federal law protects the privacy of your answers and requires us to keep all of your answers confidential. Any data that you provide will only be used by authorized personnel for statistical purposes according to the Confidential Information Protection and Statistical Efficiency Act of 2002. The only exception to this promise of confidentiality is if you tell the interviewer that you intend to seriously harm yourself or someone else; in this situation RTI may need to notify a mental health professional or other authorities.

The interview will be conducted over the phone and takes on average an hour to complete. Your participation is voluntary. You may consider some of the questions to be sensitive in nature and some of the questions also may make you feel certain emotions, such as sadness. Remember that you can refuse to answer any questions that you do not want to answer, and you can stop the interview at any time. If you become upset at any time during the interview and wish to speak to a mental health professional about how you are feeling, the interviewer can again provide you with the toll-free hotline numbers that are printed on your payment receipt from the first interview. **If you agree to complete the interview, you will receive $30 today.**

If you have questions about the study, call the Project Representative at . If you have questions about your rights as a study participant, call RTI's Office of Research Protection at (a toll-free number). You can also visit our project Website: for more information.

Thank you for your cooperation and time.

, Project Officer
Center for Behavioral Health Statistics and Quality
Substance Abuse and Mental Health Services Administration (SAMHSA)
U.S. Public Health Service
Department of Health and Human Services

Exhibit 4.3 Answering Questions About the Follow-up Interview

Respondent Question:	Suggested Response:
How long will the interview take?	The time varies, but it generally takes about an hour. Of course each person may take a little more or less time, depending on the individual. If now is not a good time, we can schedule the interview for a more convenient time.
The interview is too long!	The time varies depending on the individual. We can always stop and finish later if necessary. OR If now is not a good time, we can schedule the interview for a more convenient time.
Will my name be associated with the interview?	Absolutely not. I will not record your name on any documents. I needed your name to contact you about the interview, but that information is kept separate from your responses.
Are the questions personal?	Some questions may seem personal to some people. All answers are confidential and you don't have to answer any questions that you don't feel comfortable answering.
What type of questions will you ask? What types of personal information will you be asking?	You will be asked questions about various mental health issues such as mood, anxiety, and substance dependence.
How do I refuse to answer a question?	Simply tell me that you do not want to answer that question.
Why am I being chosen for the interview?	You were randomly chosen for this special study, which asks questions about various mental health issues. Knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.
What is your background? Why a different interviewer?	I have received special training to conduct the interviewing process for this study.
Who do you work for?	I work with Research Triangle Institute, a private, nonprofit research organization. RTI is conducting this study for the U.S. Public Health Service.
Are you going to call the police?	If you tell me that you intend to seriously harm yourself or someone else, I may need to notify the appropriate authorities.

This page intentionally left blank.

5. CONDUCTING THE INTERVIEW

5.1 Introduction

Chapter 4 described preparing for an interview, including the informed consent process and permission to audio record the interview. This chapter reviews the actual interview process including information on administering the SCID, scoring the components, and the procedures to follow at the close of the interview. Also covered are suggestions for dealing with some respondent situations you may encounter, including respondents in distress.

5.2 Conduct during the Interview

Recall your involvement on this project is for research purposes only, and you MUST NOT offer psychiatric or medical counseling, advice, diagnosis, treatment or other mental health services, even if you are qualified to do so. Refer to ***Section 2.5*** for details on this important protocol.

5.3 SCID Overview

The follow-up interview is a slightly modified version of the Structured Clinical Interview for DSM-IV for Axis I Disorders (SCID-I), which is a semi-structured interview to assess various mental disorders according to the criteria set forth by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV). A semi-structured interview is one in which a sequence of structured questions and follow-up "probes" are carefully chosen in advance to ensure consistency across respondents. Targeted, consistent questions help turn subjective responses into objective data that can be readily compared. At the same time, a properly administered interview extends well beyond the written probes provided in the protocol. As a clinical interviewer, you will be required to use your research and clinical skills as well as judgment to determine when you have gathered enough information to confidently assess the presence or absence of a symptom. If more information is needed, you should probe further until you can confidently assess whether the symptom is present or absent. As is true with any assessment, establishing rapport and maintaining a non-judgmental, objective stance are also critical behaviors to gathering good, valid information.

5.3.1 SCID Content

The Mental Health Study version of the SCID consists of the following diagnostic modules:

A. Mood Disorders

- Major Depressive Episode – Past Year
- Major Depressive Episode - Lifetime
- Manic Episode – Past Year
- Manic Episode – Lifetime
- Dysthymic Disorder – Recent

B/C. Psychotic Symptom Screen – Past Year

D. Mood Differential

E. Anxiety Disorders

Posttraumatic Stress Disorder – Past Year
Panic Disorder – Past Year
Panic Disorder with Agoraphobia – Past Year
Agoraphobia without History of Panic Disorder – Past Year
Social Phobia – Past Year
Specific Phobia – Past Year
Obsessive Compulsive Disorder- Past Year
Generalized Anxiety Disorder – Past Year

F. Eating Disorders

Anorexia Nervosa – Past Year
Bulimia Nervosa – Past Year

G. Intermittent Explosive Disorder – Past Year

H. Substance Use Disorders

Alcohol Dependence – Past Year
Alcohol Abuse – Past Year
Non-alcohol Substance Dependence – Past Year
Non-alcohol Substance Abuse – Past Year

J. Adjustment Disorder – Past Year

Specifications for each of these SCID modules can be found in ***Appendix D*** of this handbook.

These modules are preceded by the Introduction, Overview, and Screening sections and followed by the End of Interview and Interviewer Debriefing sections. The back of the SCID booklet also contains supplemental questions for discerning whether a mood or anxiety disorder has been induced by a general medical condition or substance use, as well as the Distressed Respondent and Cognitive Impairment Protocols discussed later in this chapter.

5.3.2 SCID Format

The diagnostic modules of the SCID have a three column format. The column on the left side of each page contains the interview questions and structured probes that help determine whether or not a criterion is met. The actual DSM-IV diagnostic criterion is printed in the center column and the codes for scoring the symptom are in the right column. ***Exhibit 5.1*** contains a sample SCID page.

Exhibit 5.1 Sample SCID Page

SCID-I (for DSM-IV-TR) *Anorexia Nervosa Past Year (March 2011)* *Eating Disorders* *F.1*

F. EATING DISORDERS

ANOREXIA NERVOSA **ANOREXIA NERVOSA CRITERIA**

IF SCREENING QUESTION #8 EQUALS 1, CIRCLE 1 AND GO TO ***BULIMIA*** ON PAGE F.3

IF SCREENING QUESTION #8 EQUALS 2 OR 3, CIRCLE 3 AND CONTINUE: You've said that there was a time in the past year When you weighed much less than other people thought you ought to weigh . . .

SCREEN Q#8	
NO	YES
1	3

F1

Why was that? How much did you weigh? How tall are you?

_____

_____

_____

A. Refusal to maintain body weight at or above a minimally normal weight for age and height (e.g., weight loss leading to maintenance of body weight less than 85% of that expected; or failure to make expected weight gain during period of growth, leading to body weight less than 85% of that expected)

? 1 2 3 F2

GO TO ***BULIMIA***, F.3

At that time, were you very afraid that you could become fat?

_____

_____

B. Intense fear of gaining weight or becoming fat, even though underweight.

? 1 2 3 F3

GO TO ***BULIMIA***, F.3

At your lowest weight, did you still feel too fat or that part of your body was too fat?

IF NO: Did you need to be very thin in order to feel good about yourself?

IF NO AND LOW WEIGHT IS MEDICALLY SERIOUS: When you were that thin, did anybody tell you it could be dangerous to your health to be that thin? (What did you think?)

_____

_____

_____

C. Disturbance in the way in which one's body weight or shape is experienced; undue influence of body weight or shape on self-evaluation, or denial of the seriousness of the current low body weight

? 1 2 3 F4

GO TO ***BULIMIA***, F.3

?=inadequate information *1=absent or false* *2=subthreshold* *3=threshold or true*

5.4 Administering the SCID

5.4.1 Asking the Questions

The questions printed in the SCID are to be asked ***VERBATIM AND IN THE ORDER PROVIDED***, unless the respondent has already clearly provided **all** of the necessary information earlier in the interview process. In such instances, you should confirm the information already obtained by paraphrasing the original question. Do not just assume that the symptom is present without asking for confirmation, because some aspect of the criterion may not have been adequately explored. However, you SHOULD NOT rely solely on the questions provided in the protocol. Further probes are almost always necessary to determine, with confidence, whether a symptom is present, and if so, if it is clinically significant (meets threshold). Follow-up questions should be phrased in ways that are understandable for respondents based on their age and cultural background. ***Ask as many follow-up questions as are necessary to rate each symptom with a high degree of confidence.***

5.4.2 Common Pitfalls to Avoid

When interviewing, be sure to avoid the following:

- Asking leading questions (e.g., "So, this hasn't really bothered you, right?")
- Failure to listen closely (e.g., not probing about behaviors/experiences/symptoms previously reported)
- Not providing adequate documentation (e.g., recording "Yes" on the SCID without any supporting remarks, examples, etc.)
- Inadequate probing (e.g., administering the interview as if it had a "True-False" format)
- Asking vague and/or inappropriate questions (e.g., "Have you ever been in trouble due to your drinking?")

5.4.3 If You Feel the Respondent Is Not Providing Valid Data

What you should do if you feel that a respondent is not providing good, valid data depends on the nature and source of your concern:

- ***Poor Comprehension*** – Some respondents have difficulties understanding the questions or the underlying constructs or are not inclined to admit the symptom exists. Your first step will be to rephrase or explain the target symptom in words that the respondent might better understand.
- ***Poor Insight and/or Response Biases*** – Poor insight and response biases are also common obstacles to gaining good, valid information. Given the sensitive nature of these questions, the validity of the data is under constant threat by under-reporting, positive impression management, and/or denial.

  Typically, this difficulty arises with the respondent reporting a heavy pattern of substance use but denying negative consequences of his/her use. However, an occasional respondent reports a relatively light pattern of use but then over-endorses or pathologizes the substance use.

In all cases, if the respondent is reporting information that is inconsistent with his/her previous reports and/or your clinical judgment, gently point out the discrepancy. Try using phases like this:

"I don't think I'm following you…"

"Please help me understand this…"

"So let me make sure I've got this right…"

It can be difficult to maintain rapport when you are pointing out discrepancies. But in the name of science, please err on the side of being slightly confrontational to increase confidence in the validity of the data.

5.5 Scoring the Symptoms

Each criterion symptom is coded as "?," "1," "2," or "3." Use your clinical judgment to rate each symptom based on the respondent's verbatim answers and follow-up questions.

1 = Absent or false – The symptom is absent or false.

2 = Subthreshold – The threshold for the criterion is almost, but not quite, met. A "2" rating is usually given when the behavior in question occurred, but at a frequency, duration, or intensity less than that required to meet full threshold. Code a "2" if the behavior specified was present in the past 12 months, but not at a clinically significant level.

3 = Threshold or true – The criterion is clearly met. A "3" rating is given if the behavior was present at a clinically significant level during the past 12 months. In order to be coded as a "3," a symptom needs to be clinically significant (i.e., problematic and/or distressing), and in most cases, also has to be persistent or recurrent.

? = Inadequate information – A question mark ("?") is a **temporary code** and should be coded when you are unable to determine after probing whether the criterion has actually been met. We strongly discourage the use of this code, but if you truly cannot determine the code at that time, code "?" and continue on with the interview. If, in the course of the rest of the interview, it becomes clear how the criterion should have been coded, please go back and code the correct category and note on the appropriate lines what information helped you determine the code. If at the end of the interview it is still unclear whether the criterion has been met, go back to the question and try one more time to clarify. If it is still unclear after additional probing, code the lower of the possible codes. For example, if you are not sure if it should be a "2" or a "3," code as a "2."

If you are not certain how to code a particular item or are feeling unsure about whether you have gathered enough information, collect as much information as you can and then consult with one of the clinical supervisors for her/his opinion. Call preferably within an hour of completing the interview. The CS will be happy to talk with you about the coding decision and provide guidance/feedback on proper SCID administration procedures.

Exhibit 5.2 contains a list of "Dos and Don'ts" for the administering the SCID.

Exhibit 5.2 Dos and Don'ts of SCID Administration

DO	DON'T
Emphasize you are using a standardized interview protocol and are required to ask all questions verbatim.	Apologize for the questions you are asking, clarifying responses, or the length of the interview.
Stick to the initial questions, as they are written, except for minor modifications to consider what the respondent has already said, or to request elaboration or clarification.	Make up your own initial questions because you feel that you have a better way of getting the same information. Your minor "improvement" may have a major unwanted effect on the meaning of the question.
Focus on obtaining the necessary information to judge all of the particulars of the criterion under consideration. Ask additional clarifying questions to elicit details in the respondent's own words, such as "Can you tell me about that?" or "What do you mean by that?"	Focus only on getting a "Yes" or "No" answer to the question. It often takes additional questions and probing to determine whether or not the criterion has been met during the past 12 months.
Use your judgment about a symptom, consider all of the information available, and gently confront the respondent about responses that conflict with other information.	Automatically accept a respondent's response if it contradicts other information or you believe it is not valid.
Make sure the respondent understands what you are asking. It may be necessary to repeat or rephrase questions or to ask if he/she understands. In some cases it may be valuable to describe the entire symptom cluster you are asking about (e.g., withdrawal). Follow the lead of the respondent in the terms you use to refer to a substance. For example, if he/she uses the term "pot," then you do the same.	Use words or jargon that the respondent does not understand (e.g., referring to low-level chronic depression as "dysthymia")
Always be sure that you and the respondent are focusing on the same time period for each question.	Assume that the respondent is remaining cognizant of the timeframe.
Take care to communicate that you are listening to the respondent. For example, "You told me you got a DUI last month. During the past 12 months, has there been another time in which you were in trouble with the law because of your alcohol use?"	Ignore information that you have already obtained.

5.6 Concluding the Clinical Interview

At the conclusion of the SCID, closing statements must be read to each respondent to ensure all necessary information is provided. The End of Interview Script (***Exhibit 5.3***) thanks the respondent for his or her time and mentions the possibility of needing to speak with a counselor. Remember that during the initial NSDUH interview the respondent received the $30 cash payment from the field interviewer in appreciation of the time spent completing this telephone interview. ***Exhibit 5.4*** contains a copy of the Follow-up Interview Payment Receipt provided by the interviewer to the respondent. Note the contact information for both the Lifeline and treatment hotlines. As the script indicates, provide the Lifeline number for an interested respondent who does not have the receipt.

As you finish the End of Interview script, ask for any further questions and then conclude the call. After ending the call, click the red "Stop Recording" button. Complete the Interviewer Debriefing questions within the SCID (see ***Appendix D*** for details).

Be sure to document the results of your contact, as explained in the next chapter.

Except while editing the document prior to shipment, be sure to protect the confidentiality of the paper SCID by keeping it in a locked location at all times.

Exhibit 5.3 End of Interview Script

That was my last question. Thank you for your time and cooperation in completing this interview.

Sometimes the personal issues we've discussed cause people to become upset and in need of speaking with a counselor. If you are feeling upset or disturbed by the personal issues we have discussed in this interview and would like to talk with someone about your feelings, we suggest you call your doctor, counselor, or other treatment provider if you are currently under someone's care. If not, there is also a national lifeline number you can call. This number is on the receipt for the $30 you received for this interview from the interviewer who met with you earlier. Do you still have that receipt?

IF NO: We would like to give you the hotline number for the National Lifeline Network where counselors are available to talk at any time of the day or night. They can also give you information about (additional) mental health services in your area. Their toll-free number is 1-800-273-8255.

IF YES: OK. Please know that counselors at the National Lifeline are available to talk at any time of the day or night. They can also give you information about mental health services in your area if you request this information.

Do you have any additional questions you'd like to ask me before we end our call?

Thank you again for your time, and have a good (day/afternoon/evening).

Exhibit 5.4 Follow-up Interview Payment Receipt

Follow-up Interview Payment Receipt

United States Public Health Service

and

Research Triangle Institute

thank you for agreeing to participate in a special study for the 2011 National Survey on Drug Use and Health.

In appreciation of your participation in this important study, you are eligible to receive a $30 cash payment.

Since maintaining the confidentiality of your information is important to us, your name will not be entered on this form. However, the interviewer must sign and date this form to certify you received (or declined) the cash payment.

_____	_____	__ __ __ __ __ __ __ __ __ __ - __
Interviewer	Date	Case ID
☐ Accepted Cash Payment		☐ Declined Cash Payment

If you ever feel that you need to talk to someone about mental health issues, you can call the National Lifeline Network. Counselors are available to talk at any time of the day or night and they can give you information about services in your area.

1-800-273-TALK or 1-800-273-8255
1-888-628-9454 (Spanish)
http://suicidepreventionlifeline.org/

If you ever feel that you need to talk to someone about drug use issues, you can call the Center for Substance Abuse Treatment's national referral service. This is a 24-hour service that will help you locate treatment options near you.

1-800-662-HELP or 1-800-662-4357
1-800-487-4889 (TDD)
1-877-767-8432 (Spanish)
http://findtreatment.samhsa.gov

Disposition: Top copy to Respondent, yellow to Field Supervisor, pink to Field Interviewer.

5.7 Handling Unusual Situations

In most cases the interview progresses smoothly as planned. For the cases that do not, the following explanations should help you properly handle the case.

5.7.1 Breakoff

You may start the interview but the respondent cannot complete the interview at that time. Perhaps a child runs in with an injury, or the respondent forgot another scheduled event and must leave. Or, perhaps the respondent seems distracted by argumentative voices in the background, in which case you should diplomatically end the interview ("It sounds like this is not a good time to talk. I can call back later."). If you suspect the respondent is cognitively impaired, follow the protocol in ***Section 5.7.3***.

If possible with a breakoff, try to schedule a convenient time for the respondent to complete the interview. Record the details of the appointment and the situation in the CMS. Try to complete the case in the next day or two to keep the portions of the interview close together.

When you begin the interview again, you must complete the informed consent requirements again by reading the complete introductory script (refer to ***Section 4.4***).

5.7.2 Refusal

At times a respondent may refuse to begin the interview, or may breakoff the interview in an unfriendly manner. In these refusal cases, tactfully try to persuade the respondent to begin or continue. Be sure to listen carefully to the respondent's concerns so that you can first acknowledge the concern or misunderstanding, and then address it by providing appropriate additional information.

Use either the pending breakoff or the pending refusal code for the case. Be sure to discuss the case with your Data Collection Manager. There are currently no plans to recontact refusals for the Mental Health Study, but it is still important to discuss the case and note the issues before finalizing the case.

5.7.3 Cognitive Impairment Issues

Occasionally as you talk with a respondent, you may suspect that the respondent is intoxicated or otherwise impaired and currently unable to complete the interview.

<u>If you have started the SCID</u>, stop and administer the Short Blessed Scale. These questions, shown in ***Exhibit 5.5***, are found in the Cognitive Impairment Protocol Section at the back of the SCID Booklet.

- For 10 or less errors, resume the interview. Be sure to note the situation in the interviewer debriefing questions at the end of the SCID.
- For more than 10 errors, breakoff the interview, documenting the situation when you enter the status code. Should you be able to complete the interview at another time, include the reason for the breakoff in the debriefing questions. Be sure to review the portions of the SCID completed earlier, to verify accuracy, before continuing.

<u>If you have not yet started the actual SCID</u> but are still completing the introduction and informed consent process when you suspect the respondent is cognitively impaired, breakoff the contact as explained in ***Section 5.7.1***.

Exhibit 5.5 Short Blessed Scale Exam

SHORT BLESSED SCALE EXAM

THE SHORT BLESSED SCALE IS TO BE COMPLETED AT ANY POINT DURING THE INTERVIEW IF THE RESPONDENT APPEARS TO BE COGNITIVELY IMPAIRED.

ERROR SCORES

SB-1. What year is it now? _____

CIRCLE 4 FOR ANY ERROR 0 4

SB-2. What month is it now? _____

CIRCLE 3 FOR ANY ERROR 0 3

Please repeat this phrase after me: John Brown, 42 Market Street, Chicago.

NO SCORE – FOR ITEM SB-6.

SB-3. About what time is it? _____

CIRCLE 3 FOR ANY ERROR 0 3

SB-4. Please count backwards from 20 to 1.
[20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, 1]

2 PER ERROR 0 2 4

SB-5. Please say the months of the year in reverse order.
[DEC, NOV, OCT, SEP, AUG, JUL, JUN, MAY, APR, MAR, FEB, JAN]

2 PER ERROR 0 2 4

SB-6. Please repeat the phrase I asked you to repeat before.
[JOHN BROWN/ 42 MARKET STREET/ CHICAGO]

2 PER ERROR 0 2 4 6 8 10

TOTAL NUMBER OF ERRORS IN SB-1 TO SB-6: ..... _____

IF THE TOTAL NUMBER OF ERRORS IS GREATER THAN 10, TERMINATE THE INTERVIEW.

5.7.4 Distressed Respondents

Due to the sensitive nature of portions of the SCID interview, you may encounter a respondent who becomes upset during the interview. In situations where you feel that based on indirect and direct statements made by the respondent that he or she has had suicidal or homicidal thoughts within the past two weeks or is otherwise distressed, use the Distressed Respondent Protocol (see ***Exhibit 5.6***) to determine the proper course to follow. So that it is always readily available, this Protocol is printed at the back of each SCID booklet. The first page of the protocol lists issues to be aware of before, during, and following the interview as you interact with the respondent. Although you may be qualified, you are NOT to provide counseling or other services. For your work on this study, you must follow the approved protocol.

The chart at the bottom of the first page lists various scenarios you may encounter and refers to the more detailed instructions on the following pages. Deciding which scenario to use is a judgment call for you to make based on your interactions with the respondent.

In certain scenarios and in ANY case where you are unsure of the proper scenario or correct approach, you are to consult with project staff to determine the best course of action for the situation. The Clinical Supervisors ( ), who are licensed psychologists, will act primarily as a sounding board for you. If there is a question about what action to take in response to your interactions with a respondent, you will discuss the situation with one of the Clinical Supervisors, and the Clinical Supervisor will make the final decision as to what action, if any, should be taken beyond documenting the situation. These steps are for your protection and are required by RTI's IRB.

In accordance with the Distressed Respondent Protocol or after consulting with one of the Clinical Supervisors ( ), you may need to call an emergency psychiatric care provider or 911 in the respondent's area (e.g., crisis center, inpatient hospital) so emergency help can be dispatched to the respondent's location as needed. Obtain referral information for the respondent's area from the SAMHSA website (http://mentalhealth.samhsa.gov/databases/) and the national 911 database.

When calling an emergency psychiatric care provider, use 3-way calling to keep the respondent on the line while you call. Be familiar with and practice using 3-way calling on your telephone prior to initiating any interviews so you can perform this operation smoothly, if needed. Note: with 3-way calling, once the respondent is speaking with the other person, you must **remain on the line** in order to maintain their connection.

Once speaking with an emergency psychiatric care representative or 911 in the respondent's area, use the script in the Distressed Respondent Protocol to explain the situation, and provide the respondent's name, telephone number, and, if needed, the address reported by the respondent (if known) as well as the respondent's home address (if the respondent reports a different location). To protect respondent confidentiality, when possible, only share information required to properly assist the respondent. The emergency care representative or 911dispatcher will need the entire address to send assistance.

To report the respondent's address to the emergency psychiatric care representative or 911 dispatcher:

- ask the respondent for the address of his/her current location; and
- obtain the respondent's home address from the CMS. From the CMS, access the case (through the Case Status link), and click as directed in the Address column to display the address for the QuestID. To protect confidentiality the address is not provided as assignment information, but is readily available should you need it in an urgent situation.

While you are not to participate in the conversation between the respondent and the emergency care representative, do pay attention to be sure the respondent receives instructions for how to receive emergency services. If the emergency care representative is not able to provide services to protect the safety of the respondent, call 911 in the respondent's area. Note the protocol includes these instructions to ensure proper handling of each respondent's situation.

When finished with any distressed respondent situation, be sure to document both the situation and your response by completing an online Incident Form accessible through the Main Page of the CMS. See ***Exhibit 5.7*** for the types of incident information to enter. **Proper documentation of all distressed respondent situations is critically important** to protect respondents' rights. For non-emergency situations (Scenarios 1, 3, 5), contact to consult with them about the course of action that may need to be taken to protect the respondent or other possible victims. If you have questions about what to include in an Incident Report, or if you want to debrief about the case, you may also call . Additionally, you must contact immediately for any cases for which imminent danger is or may possibly be involved (Scenarios 2 and 4) to receive supervision and debrief. If it is an urgent situation and is unavailable within 15 minutes, try the next person on the list, and so on. In addition to the study contact summary printed in the inside cover of this handbook, the chart below lists these contacts:

Call:	Cell Phone Numbers:

Please note that do not have clinical training. Calling one of them is important to alert project staff that the Distressed Respondent Protocol has been enacted and you are attempting to contact the Clinical Supervisors ( ).

The study's IRB approval rests on following these carefully developed procedures exactly.

Exhibit 5.6 Distressed Respondent Protocol

Specific Guidelines

If respondents report any of the issues listed below during any interactions with the recruiter or clinical interviewer, including before, during, or after a telephone screening or interview, the staff member will immediately refer to the scenario chart below and follow the instructions provided. Details of all incidents will be documented on the case management system and reported to project management staff immediately.

- Has had **any suicidal thoughts in the past two weeks (p. A.3)**, including
  - passive suicidal thoughts (i.e. thoughts or wishes about his/her death **in the absence of** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, or intention of dying or attempting suicide) **[SCENARIO 1]** or
  - active suicidal thoughts (i.e. thoughts or wishes about his/her death **combined with** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, the intention of dying or attempting suicide, and the means to carry out that plan) **[SCENARIO 2]**

- Has had **any homicidal thoughts in the past two weeks**, including
  - passive homicidal thoughts (i.e. thoughts or wishes about seriously harming someone else **in the absence of** thoughts about specific ways in which s/he could seriously harm another person, plans for how s/he could seriously harm another person, intentions of seriously harming another person) **[SCENARIO 3]** or
  - active homicidal thoughts (i.e. thoughts or wishes about seriously harming someone else **combined with** thoughts about specific ways s/he could seriously harm another person, plans for how s/he could seriously harm another person, the intention of seriously harming another person, and the means to carry out that plan) **[SCENARIO 4]**

Scenario Chart		
Scenario Number	**Individual at Risk of Harm**	**Imminent Danger?**
1	Self	No
2	Self	Possible / Yes
3	Other(s)	No
4	Other(s)	Possible / Yes
5	No risk of harm; respondent is agitated or upset	No

Exhibit 5.6 Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
1	Self	No

STEPS

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about your death or dying. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about death and dying. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL IF YOU HAVE QUESTIONS OR WOULD LIKE TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

Exhibit 5.6 Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
2	Self	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you are thinking about harming yourself. I strongly suggest that we contact emergency care services in your area, such as a crisis center or nearby hospital. I am going to look-up that number. Can you remain on the line while I do that? It may take a few minutes.

> **IF NO:** Okay, if I don't connect you with the local emergency care provider, then I will need to call the provider myself to see if they can send someone to you who can provide the care you need in order to keep you safe. I'll call you back to let you know what I find out.

B. FIND THE NEAREST EMERGENCY PSYCHIATRIC SERVICES USING THE SAMHSA WEBSITE (http://mentalhealth.samhsa.gov/databases/). **SEARCH FOR INPATIENT MH TREATMENT USING THE R'S CURRENT ZIP CODE.**

C. CALL THEIR LOCAL INPATIENT PSYCHIATRIC CARE FACILITY OR CRISIS CENTER AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming (himself/herself) and I am concerned about (his/her) safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

> **IF ASKED FOR NSDUH OVERVIEW:** This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (his/her) well-being, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

D. GIVE R FIRST NAME, TELEPHONE NUMBER, AND ADDRESS (IF KNOWN) TO LOCAL EMERGENCY CARE REPRESENTATIVE. IF THEY ARE UNABLE TO PROVIDE SERVICES THAT ENSURE THE R'S SAFETY, SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

E. IF R ON THE OTHER LINE, CONNECT R TO EMERGENCY CARE REPRESENTATIVE OR LOCAL 911 DISPATCHER AND STAY ON THE LINE; IF YOU HANG-UP, THEIR CONNECTION WILL ALSO END.

IF R NOT ON THE OTHER LINE, END CALL WITH THE EMERGENCY CARE PROVIDER OR LOCAL 911 DISPATCHER AND ATTEMPT TO CONTACT R AGAIN WITH AN UPDATE.

Exhibit 5.6 Distressed Respondent Protocol (continued)

SCENARIO 2 (Continued)
F. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Exhibit 5.6 Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
3	Other(s)	No
STEPS		

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about seriously harming someone else. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about seriously harming someone else. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF DIRECTED BY ONE OF THEM, FOLLOW SCENARIO 4 FOR POSSIBLE IMMINENT DANGER TO OTHERS. FILL OUT ONLINE INCIDENT REPORT.

1.1.1

Exhibit 5.6 Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
4	Other(s)	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND END CALL.

B. SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

C. CALL THEIR LOCAL 911, AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming another individual. I am concerned about this individual's safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

> **IF ASKED FOR NSDUH OVERVIEW:** This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (him/her) harming someone else, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

D. GIVE R FIRST NAME, TELEPHONE NUMBER, ADDRESS (IF KNOWN), AND VICTIM'S IDENTIFYING INFORMATION TO LOCAL 911 DISPATCHER. END CALL.

E. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Exhibit 5.6 Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
5	No risk of harm; respondent is agitated or upset	No

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: I know these questions are very personal, and they seem to be upsetting you. Do you have a doctor or someone you can talk to about how you are feeling?

IF YES: I suggest that you call that individual immediately so that she or he can help you talk about and work through how you are feeling. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline is a 24-hour hotline that you could call to discuss this with a counselor. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away or call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL IF YOU HAVE ANY QUESTIONS OR NEED TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

Exhibit 5.7 Incident Report Form

Incident Report Form

When completing the Incident Report Form, DO NOT include identifying information such as the respondent's name or address
*required fields

Project Title:	National Survey on Drug Use and Health (NSDUH): Mental Health Surveillance Study
***Date of Incident:**	4 / 15 / 2011
***Respondent's Location:**	

***Participant's QuestID :**

***Describe in detail the nature and timing of this incident:** 6000 characters remaining

If Report or Referral was made to an outside agency:

Date and time of report/referral: MM/DD/YYYY / /
HH:MM AM/PM : N/A

Agency and person who took report/referral:

Current disposition of situation: 1000 characters remaining

Save | Submit Report | Reset | Cancel

5.7.5 Unable to Contact

If no one answers your call, be sure to document that attempted contact in the CMS. Plan your next call at a different time of day or on another day of the week.

If you do speak with someone but not the actual respondent, try to determine a good time to catch the respondent at home. If that person on the phone offers another telephone number such as a cell phone or a work number as a way to reach the respondent, make note of the number. Do not **ask** for an alternate number, but take the information if provided. While you would not want to complete the interview while the respondent is working, you could use the work number to contact him or her about setting an appointment for the interview.

For cases with ongoing challenges, be sure to discuss the situation with your Data Collection Manager.

5.7.6 Other Non-Response

Rarely, you may encounter a person who should not have been included as a potential study respondent (youth under 18, language barrier, or physical impairment preventing a telephone interview). In such a case, select the appropriate code and describe the situation in the CMS. During your next conference call, be sure to let your Data Collection Manager know about the case.

6. DOCUMENTING RESULTS

6.1 Introduction

The documentation and reporting of interviewing activities you provide is used to closely monitor the ongoing data collection efforts. There are a number of ways you document and report your activities, through:

- Record of Calls. The actions you take to complete interviews are documented in the Record of Calls on the CMS (see ***Section 6.2)***.
- Audio Transfer. Upload the audio recording of each interview within 2 hours of its completion (see ***Section 3.6***).
- SCID Shipment. FedEx edited SCID booklets within 48 hours of completing the interview (see ***Section 6.3***).
- Conference Calls. You and your Data Collection Manager will have scheduled conference calls usually once each week (see ***Section 7.3***).
- ePTE. Each week you will submit your electronic Production, Time and Expense report (ePTE). In the event of computer problems, paper PT&E reports serve as a backup and must also be submitted weekly (see ***Chapter 7***).

6.2 Documenting Results

Accurate and timely documentation of the final results for each case is crucial to the success of the study.

6.2.1 Record of Calls

Each case will have a "Record of Calls" associated with it which describes the details of what happened from the time the follow-up case was generated to the time it is completed. The listing will help you keep track of what happens each time you contact a respondent. It is also used by your supervisors and other RTI project staff to understand how long it takes to complete the interview process and potential problems that may arise in the process of completing those interviews.

You will track all calls in the web-based CMS from which you receive your initial assignment of cases.

6.2.2 Status Codes

The Record of Calls is essentially a listing of various "Status Codes" which you assign each time you contact a respondent regarding the follow-up interview. These status codes identify what happened during each contact or attempted contact. In addition to assigning the status code, you will enter any notes regarding what transpired. The system associates the entry date and time with the record as well. There are two types of codes: Pending and Final.

PENDING CODES

Pending codes are used when the case is not yet complete. For cases to which you assign pending codes for breakoffs, refusals, or other, you should discuss the situation with your Data Collection Manager who will help you to determine the next step. Explanations of when to use a particular code, and what action to take to resolve the situation so that the case can be completed, are provided in ***Exhibit 6.1.***

FINAL CODES

Final interview result codes indicate the case is finished, either because the interview was successfully completed or because you were unable to obtain an interview (see ***Exhibit 6.2***).

6.2.3 Recording Comments

Since it is important that project staff understand the entire process by which interviews get completed and any problems encountered while completing the interviews, please enter detailed notes. When you have contacted a respondent and scheduled an appointment to complete the interview, record the date and time of the appointment in the Notes. If you have to breakoff an interview, record notes about why the breakoff occurred as well as any information about re-contacting the respondent. If the respondent refuses to complete the interview, or refuses to allow you to record the interview, please document the reasons for refusing.

Exhibit 6.1 Pending Status Codes

Status Code	Status Code Description	Use This Code When...	Follow Up Actions to Take
50	**APPOINTMENT FOR FOLLOW-UP INTERVIEW**	The respondent cannot complete the interview now but you have scheduled an appointment	• Record the time and date of the appointment in the Notes. • If you become unavailable to call at the time of the appointment, be sure to inform your Data Collection Manager so the case can be assigned to another interviewer.
51	**R UNAVAILABLE**	Spoke with someone but not respondent	• Ask for a good time to reach the respondent and record in the Notes.
52	**RING NO ANSWER, BUSY**	The phone rings continually or is busy	• Complete the ROC, making a note if busy
53	**MACHINE/VOICE MAIL: NO MSG**	An answering machine or other voice mail responds and you did not leave a message	• Complete the ROC, making notes • After the initial call with a message, leave a message every 3rd time you encounter an answering machine.
54	**MACHINE/VOICE MAIL: LEFT A MSG**	An answering machine or other voice mail responds and you left a message	• Complete the ROC, making notes. • If you continue to have trouble reaching the respondent, consider leaving a message with your number so the R can call you back to arrange an appointment.
56	**BREAKOFF (PARTIAL INTERVIEW)**	You have started the interview, but must stop the call	• Arrange another time to contact the R and complete the interview. • If the breakoff is unfriendly (R refuses to continue), tactfully try to persuade the respondent to continue. • Record the reason for the breakoff in the "Notes" section.
57	**REFUSAL**	The respondent refuses to let you start the interview	• Tactfully try to persuade the respondent to participate. • Record the refusal reason in the Notes. • Contact your Data Collection Manager and discuss the case.
59	**OTHER, SPECIFY**	Use this code for cases that do not fit any of the other categories	• Be sure to fully describe the situation in the Notes section • Discuss with your Data Collection Manager

Exhibit 6.2 Final Status Codes

Status Code	Status Code Description	Use This Code in This Situation
70	**INTERVIEW COMPLETED, AUDIO RECORDED**	The respondent has completed the follow-up interview AND you were able to record the interview. This is the BEST result code and is the desired result for all selected interviews.
71	**INTERVIEW COMPLETED, NO AUDIO (REFUSAL)**	The respondent has completed the follow-up interview, but refused to allow the audio recording of the interview. In the Notes, specify why the audio was refused, if known.
72	**INTERVIEW COMPLETED, NO AUDIO (TECHNICAL PROBLEM)**	The respondent has completed the follow-up interview, but the audio recording of the interview did not take place due to technical problems. In the Notes, specify why the audio could not be recorded, if known. Work with your Data Collection Manager and Technical Support as needed to resolve any technical issues.
73	**BREAKOFF – PARTIAL INTERVIEW**	You have started the interview and the respondent either refuses to allow you to complete the interviewing process or is unable to complete the interview. Speak with your Data Collection Manager before assigning this code.
74	**UNABLE TO CONTACT RESPONDENT AFTER REPEATED CALLS**	We hope that with persistence this code will be used rarely. Although you may have spoken with other(s) at the number, you have not reached the respondent, or you may have spoken with the respondent earlier to set an appointment, but since then have been unable to reach him/her to complete the interview. Use this code after repeated attempts to contact the respondent to complete the interview have been unsuccessful. Speak with your Data Collection Manager before assigning this code.
75	**FINAL NUMBER PROBLEM**	The case cannot be completed due to problems with the telephone number provided. Describe the situation in the Notes. Speak with your Data Collection Manager before assigning this code.
76	**FINAL REFUSAL**	All attempts to convert the respondent to participate have been unsuccessful. Be sure you indicate the reason for the refusal in the Notes. Speak with your Data Collection Manager before assigning this code.
77	**FINAL OTHER, SPECIFY**	Use this code for cases that do not fit any of the other categories. Be sure to fully describe the situation in the Notes. Speak with your Data Collection Manager before assigning this code.
	Supplemental Status Code	
80	**MATERIALS SHIPMENT**	Once you have prepared the completed SCID for shipment, make another entry using Status Code 80 to indicate that you are FedExing the confidential SCID booklet. Enter the FedEx tracking number in the Notes.

6.2.4 Entering Records in the CMS

Following each call to a respondent, enter a status code and any notes into the web-based CMS so your supervisors and project staff can view the most current information regarding the status of the case.

To enter the status codes in the CMS, follow the same steps outlined in ***Section 4.2.2*** to get to the "Case Status" page and obtain a listing of all your cases.

Locate the appropriate Quest ID and click on the current status code displayed in the **Status Code** column. A pop-up box entitled **Assign a New Event/Status Code** appears.

The Quest ID displays at the top of the box along with the most recent status code associated with the case.

In the **New Event** box select the appropriate code from the drop down menu.

Record any notes associated with the contact in the "New Event Note" box. (See ***Exhibits 6.1*** and ***6.2*** for details on the codes.)

Then click the **Update** button at the bottom of the box to save the entry. The CMS automatically associates the date and time of your entry with that record, which means it is helpful to make your entries soon after the call is completed. NOTE: The CMS always uses Eastern time.

Back at the Case Status page, please note the new entry displays in the row for the case AFTER you click "Filter Data" (refer to ***Section 4.2.2***).

To view all the Status codes associated with the case (i.e. the case history, or Record of Calls) simply click on the Quest ID on the Case Status page. A pop-up box appears displaying all status codes, the time and date those entries were made and any notes entered.

6.3 Disposition of Materials

Once you complete the interview, the last step is to distribute the information and interview materials as shown in the below chart.

Item:	Send to:	Method:	Time Frame:
SCID booklet (with Cover Sheet completed) placed in manila envelope preprinted with Confidential Information Sheet	RTI	FedEx	Within 48 hours
Audio file	Secure website	Computer transfer through MH Case Manager	Within 2 hours of completing the interview

As explained in ***Section 3.6***, it is best to upload the audio file as soon after the interview as possible.

Soon after, but within 24 hours, edit the SCID. To prepare the edited SCID for shipment, complete the SCID Cover Sheet (Transmittal Record) (see ***Exhibit 6.3***) by entering:

- your ID number,
- the Quest ID for the case,
- the date the interview was completed, and
- the date you are shipping the SCID to RTI.

Once you have completed the Cover Sheet of the SCID booklet, place it in one of the provided manila envelopes with a Confidential Information Sheet (see ***Exhibit 6.4***) preprinted on the envelope. Then seal the envelope prior to placing it in the FedEx package. ***Section 7.8*** contains detailed instructions for FedEx packages.

Once you have successfully shipped the SCID, enter that information in the CMS to document the FedEx tracking number of the SCID package. Add an additional status code of 80 and **include the FedEx tracking number** in the notes. The system uses the 80 code to generate an e-mail message to your Data Collection Manager and the Clinical Supervisors so they know the package has been sent. The e-mail includes the tracking number, should there be any delivery issues with this confidential package.

6.4 Summary of the Interview Process

For each case assigned to you, there are a number of steps to complete. As you work, use the CI Case Checklist shown in *Exhibit 6.5* as a guide to be sure you have completed each step. If you have questions about procedures, first refer to the appropriate place in this handbook, then as needed, please check with your Data Collection Manager.

Exhibit 6.3 SCID Cover Sheet Transmittal Record

OMB # 0930-0110
Expiration Date: 01-31-12
V.6, 3/4/11

STRUCTURED CLINICAL INTERVIEW FOR DSM-IV AXIS I DISORDERS (SCID-I)

By

Michael B. First, M.D.; Miriam Gibbon, M.S.W.;
Robert L. Spitzer, M.D.; and Janet B. W. Williams, D.S.W.

MODIFIED BY RTI INTERNATIONAL

FOR

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

MENTAL HEALTH SURVEILLANCE STUDY

CONFIDENTIAL

SCID Transmittal Record			
Interviewer ID:		QuestID:	
		Date of Interview:	____/____/____ MM DD YY
Date Shipped to RTI:	____/____/____ MM DD YY	Date Received at RTI:	____/____/____ MM DD YY
Clinical QC by:		Date of Clinical QC:	____/____/____ MM DD YY
Edited by:		Date Edited:	____/____/____ MM DD YY

Public reporting burden for this collection of information is estimated to 60 minutes per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any aspect of this collection of information, including suggestions for reducing this burden to SAMHSA Reports Clearance Officer; Paperwork Reduction Project (0930-0110); Room 7-1045, 1 Choke Cherry Road, Rockville, MD 20857. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. The OMB control number for this project is 0930-0110.

Exhibit 6.4 Confidential Information Sheet

CONFIDENTIAL INFORMATION

IF FOUND, PLEASE CONTACT

@

Ext.

FEDEX TRACKING NUMBER: _____

Property of:
RTI International
3040 Cornwallis Road, Research Triangle Park, NC 27709

Exhibit 6.5 CI Case Checklist

Clinical Interviewer Case Checklist	
Task	**Established Deadline**
Receive Case Assignment	
Check web CMS for new cases (daily)	
Prepare for the Interview	
Gather materials	
Set up recording equipment, complete initial test recording	
Introduce and Complete the Interview	
Use Introduction script	
With permission, begin recording	
Complete SCID	
Handle distressed R situation appropriately, if needed	
When finished, thank R and stop recording	
Document and Distribute Information	
Upload audio file to web through MH Case Manager	Within 2 hours
Update case status results in CMS	Within 2 hours
Document distressed R situation, if needed	Within 2 hours
Edit SCID	Within 24 hours
Complete SCID Cover Sheet, place in manila envelope with preprinted Confidential Information Sheet, and FedEx to RTI	Within 48 hours
On CMS, document FedEx shipment by adding a code 80 and including the FedEx tracking number in the Notes	Within 48 hours

This page intentionally left blank.

7. ADMINISTRATIVE PROCEDURES

7.1 Introduction

Knowing and following all administrative procedures is vital to the success of this study. Be sure to read this chapter carefully, refer back to it when necessary, and do not hesitate to contact your Data Collection Manager whenever you have questions about administrative issues.

7.2 Terms of Employment

As a clinical interviewer, you are an employee of Headway Corporate Resources, referred to as Headway. The terms of your employment are spelled out in the Employee Guidelines and Supplemental Guidelines along with other employment forms signed by you. Headway handles the payment for hours worked and expenses, including deductions for Federal, State, and Social Security taxes. You must complete a W-4 form so that appropriate deductions are withheld.

In addition to all payroll functions, Headway is responsible for all necessary personnel documentation and the actual hiring and terminating of all employees. Your Data Collection Manager acts as your direct supervisor with full authority to manage staff.

7.3 Working with your Supervisors

During your work on this study, you will report to, and work closely with the Clinical Supervisors for issues related to the SCID and the actual interview administration process, and your Data Collection Manager for issues related to case assignments, production, contacting respondents, and administrative topics such as expenses and supplies. Your supervisors are a team and will work closely together to support you.

You will participate in regularly scheduled conference calls with your Data Collection Manager to discuss your work progress and any other important issues. More than likely you will speak with him/her at least once per week. Due to the nature of the study, it is imperative that you keep him/her informed of any problems. Administratively, your Data Collection Manager will be approving your timesheets each week so you get paid in a timely fashion.

Your contact with the Clinical Supervisors will vary, with more conferences at the beginning of data collection as you receive feedback and perfect your administration of the SCID.

As needed, refer to ***Exhibit 1.1*** for contact information for the entire MHSS management team.

7.3.1 Conference Calls

As mentioned, your Data Collection Manager will hold a regularly scheduled conference call with you usually once per week. During this call, be prepared to discuss your weekly work, and schedule of availability. In order to assign cases to you, he/she must have detailed knowledge of your schedule. Have your Case Status information from the CMS available in order to discuss any respondent contact issues you may be having.

Once a Clinical Supervisor completes the review of your work, he/she will contact you directly with any feedback. In addition to this ongoing feedback, the supervisors may periodically host group or individual conference calls to discuss important SCID administration topics.

7.3.2 E-mail

You will be issued an RTI e-mail account for use on the Mental Health Study for additional contact with your supervisors, such as documenting a difficult case or situation, and sharing any changes in your work schedule availability. Do not use this e-mail account for personal correspondence; it should only be accessed for work directly related to the Mental Health Study. Be sure to check your e-mail daily on days you have indicated you are available to work.

7.3.3 Web Updates

Timely updates of the web CMS are extremely important due to the rigid data collection schedule. The information you enter is used to generate project reports and thus affects management decisions. Delay of information entry into the CMS may cause problems in the day-to-day management of the study. Your supervisors and others at RTI need to be aware of your work as soon as possible via the web updates. Refer to ***Section 6.2*** of this handbook for detailed information on your responsibilities in updating the CMS.

7.4 Electronic Production Time and Expense (ePTE) Reports

All clinical interviewers must complete an electronic Production Time, and Expense (ePTE) report weekly and submit it in order to be paid. In addition to serving as your timesheet, the ePTE gives RTI managers a detailed summary of the tasks completed during the period covered by the report and the time and expenses required for completing those tasks. You must submit a complete and accurate ePTE report each week for all work done in order to be paid.

While you are required to submit an ePTE each week, there may be times when you are not able to access the electronic system due to computer or other problems. In such instances, you will need to submit a paper PT&E as instructed in ***Appendix B***. Keep a small supply of paper PT&Es as a back up for situations such as these. Additionally, there may be times when an ePTE is submitted and approved, and then errors are discovered, such as a forgotten expense that was not reported. You must submit a revised report using a paper report.

7.5 The ePTE System

7.5.1 Overview

The focus of this section is how to use the ePTE program. It gives detailed instructions on entering information in the various forms as well as detailing where to enter specific hours or expenses.

Similar to paper forms, the ePTE system has several screens for you to complete. Details for the various forms are given in subsequent sections.

- Default Value Setting Form – The first time you access the system, you must identify yourself and set certain values such as your RTI ID number and the project number. You will only need to complete this form the first time you use the ePTE system (see ***Section 7.5.2***).

- Entry Form – This is the form you will use as you enter the system to provide important information that defines an individual report such as the proper project number, your Data Collection Manager, and the week for which you are submitting the form (see ***Section 7.5.3***).
- Detail Form – This form has the rows and columns in which you record the details of the time spent on various tasks and a daily account of your expenses (see ***Section 7.5.4***).

In each of these forms, you can use your mouse to navigate or press the **[Tab]** key to move from one field to the next. Holding down the Shift key while pressing the Tab key (**[Shift] [Tab]**) moves the cursor backwards through the fields. The field which is currently active will have the cursor in that field or will be a highlighted button with a dotted line around it. Within the fields, use the left and right arrow keys to move around, the delete key to remove characters, and letter and number keys as needed.

Sections 7.5.5 and ***7.5.6*** provide instructions on using the system's edit checks, how to handle various messages that may appear, and tips for some advanced functions.

To access your ePTE system, double click the ePTE icon on the Desktop. Enter your RTI Employee ID number and your password when prompted.

7.5.2 Default Value Settings

The first time you enter the ePTE system, you must provide some personal identifying information. It is *critical* that you enter this information correctly since it is used to identify the reports you submit so your payment can be processed. It is particularly important that you enter your RTI ID number correctly. An example of this form is shown below.

- In the first 8 fields: Carefully enter your personal identifying information (through the phone number field). Be very careful to enter your RTI ID number correctly. Press [Tab] to move to the next field. Pressing [Shift] [Tab] moves backwards. If your name is longer than the displayed field, the program accepts the additional characters (there is a maximum of 50 characters).

- New Name/Address/Phone: When a change is made to a name, address, or phone number field (including the first time you complete the form), this box should be checked. Having the box checked means the information/changes are transmitted to Headway for processing. You must also inform your Data Collection Manager of any address changes. He/she will forward the changes to Headway so that all files are updated.

  Important: If Headway already has your correct name and address on file, press **[Space bar]** (or click Name/Address/Phone) to uncheck this box. Only leave the box checked if there have been changes to your name and/or address that Headway needs to know about.

- Supervisor Name and Area #: Next, enter the name of your Data Collection Manager, the area ID number [either 901 ( ) or 902 ( )].

- Project Number: The project number belongs in the next two fields. Initially, your default settings should be set to the corresponding training/certification project number as shown in the chart below. During data collection, each week you will divide your hours between two codes at the ratio of approximately 1/3 and 2/3 (see Dividing Charges later in this section for more information).

	Training/Certification	2011 Data Collection	Ratio
New CI	0212682.120.005	0211838.212.006	1/3
		0212682.120.006	2/3
Returning CI	0211838.212.005	0211838.212.006	1/3
		0212682.120.006	2/3

  When finished with the certification process, change the default setting to one of the two data collection numbers. The project number can be changed in the Entry Form.

- Allow Past PT&E Edit – This box is pre-filled with 15 days. Do not change this number unless instructed to do so.

- Allow PTE Type – When Data Collection is highlighted (with a box around it) press the [**Space bar**] to place a check mark in that box. At this point, leave the other boxes unchecked. If you don't make a selection here, Data Collection is automatically chosen by default.

- Save – Once you have filled in all fields and double checked all your entries, then triple check your RTI ID. Next, press [**Alt**] [**s**] (for Save) or click Save.

  If the New Name/Address/Phone field is checked when you save your information, a message box then appears asking if you want to transmit this address to Headway. Choose Yes if Headway does not have your up-to-date name, address or telephone number on file; otherwise choose No.

  If the New Name/Address/Phone field is not checked, you will not get this message box.

  After you have saved your information, the RTI ID field is shaded in light blue indicating that this field is "read-only." You can no longer change this value without obtaining a special password from your Data Collection Manager.

  NOTE: You must enter a valid RTI ID before moving on to the ePTE Entry Form.

- Help – If you have any questions about any of the fields on this form, simply press [**Alt**] [**h**] (for Help). Read the topics displayed in the help window using the **Down Arrow** key to scroll through the text. Close the help window by pressing [**Alt**] [**F4**].
- Exit – Once you have filled in all information on this form and have saved it, you can exit by pressing [**Alt**] [**x**] (for Exit) or click Exit to automatically continue with the Entry Form.

7.5.3 Entry Form

The first time you work in the ePTE system, you complete the Default Value Settings Form as described in ***Section 7.5.2***. Upon exiting that form, you see the PT&E Application Entry Form, referred to as the Entry Form. Each subsequent time you enter the ePTE system, you start at the Entry Form. An example of this form is shown below.

The default values from the Default Value Settings Form are pre-filled in their appropriate fields. In particular, note the blue shading around RTI ID and the Counting & Listing and Supervisor buttons indicating that these fields are read only. The Data Collection Manager and project number information from the Default Value Settings Form displays, but can be altered on this form.

When completed, the information on this Entry Form defines a specific ePTE record.

- Supervisor Name – Your Data Collection Manager's name is displayed. Tab through this field to prepare an ePTE for this supervisor.
- Area ID – The proper Area ID is displayed from your default values. Tab through this field to prepare an ePTE for this Area ID.
- Week Begin Date – Enter the date of the Sunday which begins the work week you are reporting. Use the format shown in the example.
- RTI Project Number – Your default project number is displayed. Tab through these fields to prepare an ePTE for this number.

  To submit an ePTE for a different project number, press delete to remove the current number and then carefully enter the correct project number and task number.

  If the new number is similar, it may be easier to edit: use the arrow keys to move to the right of the digit to be altered, press Backspace to delete and then type the new digit. If you need to alter the Default Values Form, see ***Section 7.5.6***, item 10. Please complete any changes carefully. Entering incorrect project numbers causes problems and may delay your payment.
- Travel Status – Leave the Travel Status box blank.
- PTE Type – The options available for PTE Type depend on which PTE types you marked on the Default Values Form. You can only choose from the types that are not shaded a light blue.
- Did Not Work – If you did not perform any work during the given reporting period, place a check in this box by pressing the [**Space bar**] or clicking Did Not Work. For accurate reports, it is important for you to submit a Did Not Work ePTE record by marking this box and transmitting that information. See ***Section 7.5.6***, items 6 and 7 for additional details.

There are also several buttons on the Entry Form:

- Enter PT&E – Once you have completed all fields, click the "Enter PT&E" button to proceed to the Detail Form for that report.
- Exit – If you don't want to enter ePTE information now, click the Exit button. When you exit in this way, any changes just made to the information displayed are not saved. When you return to the ePTE system, your default values appear.

  After indicating you want to exit, a confirmation box appears asking if you are sure you wish to exit. A "Yes" response takes you to the desktop. A "No" response clears the boxes and you remain at the Entry Form.
- Transmit ePT&E – Once you have finalized your entire report and marked it on the Detail Form as ready to transmit, click the "Transmit ePT&E" button to begin the process to submit your report to RTI. This process is explained more later in this chapter.

7.5.4 Detail Form

After identifying an ePTE record and choosing the "Enter PT&E" button on the Entry Form, the Detail Form shown below appears. This section explains the process of entering information on the ePTE and gives you detailed explanations for the various columns including where to enter certain hours or expenses.

Research Triangle Institute PT&E Application - Detail Form

Data Collection PT&E

Project: 0212682-120.005 RTI ID: 000999

Week Begin: 08-22-2010 Travel: N Supervisor: 321-Patience Supervisor

A		B	C									B-1
Day of Week	Date	Production	Time									Note
		# of Completed Interviews*	Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confe-rence	Other*	Daily Total*	
Sun	08-22										0.00	
Mon	08-23										0.00	
Tues	08-24										0.00	
Wed	08-25										0.00	
Thurs	08-26										0.00	
Fri	08-27										0.00	
Sat	08-28										0.00	
Totals		0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

A		D							
Day of Week	Date	Expense							
		Miles Driven	Lodging	Meals Expense*	Tele-phone*	Auto Rental*	Misc. *	Incentive*	Total Expense
Sun	08-22								0.00
Mon	08-23								0.00
Tues	08-24								0.00
Wed	08-25								0.00
Thurs	08-26								0.00
Fri	08-27								0.00
Sat	08-28								0.00
Totals		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

* Click on labels denoted by "*" for more details

Advance Repayment*

Save

Cancel

Print

How to Start

First, check that the header information at the top of the form is correct. Also make sure the dates on the left-hand side correspond to the work week for which you are reporting. If you notice any mistakes, do not enter any information. Exit this form by clicking the "Cancel" button at the top. Make the necessary corrections on the Entry Form page before proceeding.

How to Move

Within the Detail Form, move from field to field using **[Tab]**, **[Shift] [Tab]**, or any of the arrow keys. You can also use the Touchpad to click the next fields. The field that is active—that is, if you enter information it appears in that field—is shaded a light orange. On the Detail Form shown above, the active field is Sunday, Production. Once you enter an amount in a field, the left and right arrow keys allow you to move among the characters in that field. To continue to the next field, you must press **[Tab]** or **[Enter]**. Using the Touchpad to click is a simple way to move to fields in different parts of the form.

To move from one section of the form to another, use the Touchpad or press [**Tab**] from the lower right field in that section. For example, to move from "C -- Time" to "D -- Expense," use [**Tab**] or the arrow keys to move to the last cell in the Time chart, which is Saturday, Other. Press [**Tab**] to move to section D, Sunday, Miles Driven. To move back to the "C -- Time" section from the "D -- Expense" section, be in the Sunday, Miles Driven field and then press [**Shift**] [**Tab**]. The chart flows in this way:

Time → Expenses → Advance Repayment→ Notes→ Save/Cancel/Print.

How to Enter

Time entered should be rounded to quarter hours and recorded in decimal form. Do not record time in minutes. The program rounds your entries as needed, and supplies the standard two decimal place value. So, if you enter 2 hours, the program displays 2.00. If you enter 2.2 hours, the program displays 2.25. For example if you had 1 hour and 45 minutes of interviewing time, you should enter your time as "1.75." If you enter your time in minutes, "1.45," the program would automatically round ".45" to ".50". You would only get paid for one and a half hours instead of an hour and three quarters. The table below shows examples of correct and incorrect time entries:

Time Worked	Incorrect Entry	Correct Entry
1 hour and 15 minutes	1:15	**1.25 hrs.**
1 hour and 20 minutes	1:20	**1.25 hrs.**
1 hour and 30 minutes	1:30	**1.50 hrs.**
1 hour and 40 minutes	1:40	**1.75 hrs.**
1 and 3/4 hours	1:45	**1.75 hrs.**

The program automatically calculates totals for each task (column) and day (row).

What and Where to Enter

This section details the various charges and their corresponding location on the Detail Form. Please refer to these instructions and the example that follows as you are completing your reports to be sure you are making all required entries correctly.

SECTION B -- PRODUCTION:

- Enter the total number of completed **interviews** (code 70s, 71s, and 72s) each day. You must record the QuestIDs for all completed interviews and all breakoff interviews in the Notes section (see below). Make sure to keep up with your entries daily on an ePTE Working Copy (see ***Exhibit 7.1***) so you will know what you completed that day.

SECTION C -- TIME: In each column, enter the hours worked each day to the nearest .25 hour. Hours must be entered as decimals and are rounded to the nearest quarter hour. Show how your hours were spent on project work by distributing the time worked across the various columns as appropriate. The program totals your entries by row and column. Hours submitted on your ePTE should be entered in the following columns:

- Training – Except for training, this column will not be used. However, if you are authorized to spend time doing additional study, record the time in this column.
- Home Study – This column is only used in rare situations when instructed by your Data Collection Manager or Clinical Supervisor.
- Travel – Use this column for any travel required for your work.
- Contact/Locate – Record time spent contacting respondents, including time spent making appointments.
- Interviewing – Record time spent administering the SCID from start to finish including any initial conversations with the respondent during the call.
- Editing – Enter any time spent editing the SCID in this column. Do not include interviewing time here.
- Conference – Record time spent preparing for and in conference with your supervisors, compiling weekly reports, completing incident reports, and any calls with other RTI or research staff such as Technical Support.
- Other – Record time spent performing other allowable project activities such as entering information into the web as well as time spent transferring audio files and preparing FedEx packages. You must include a short description of any "Other" activity in the "Notes" section on the side of the report.

SECTION D -- EXPENSES: You must **check with your Data Collection Manager before incurring any expenses**. Expenses submitted on your ePTE should be entered in the following columns:

- Miles Driven – Record the total **miles** driven (rounded to the nearest mile) each day on project business (check to make sure that the corresponding travel time is shown in the Travel column). Mileage is reimbursed at the established government mileage reimbursement rate of $0.51 per mile. Note: This rate is set by the federal government and is subject to change.
- Lodging, Meals, Telephone, and Auto Rental – Do not use these columns unless instructed by your Data Collection Manager.
- Miscellaneous Expenses – On the ePTE, record all miscellaneous expenses under $10.00 such as reimbursement for adding 3-way calling. You must explain the expense in the Notes and receipts are not required. **Expenses over $10.00 are submitted with receipts on the Headway Expense Report** (see ***Section 7.6***) and sent to your Data Collection Manager for approval. They are NOT reported on the ePTE.

  Note that parking tickets, traffic tickets, and other fines are not reimbursable expenses.
- Incentives – For this study, you should not have any incentive expenses since respondents were paid for the telephone interview at the time of the initial interview.
- Total Expenses – The program will total your eligible expenses for each day.

Exhibit 7.1 ePTE Working Copy

Clinical Interviewer WORKING COPY of the ePTE

Data Collection ePTE	**Week Begin**	**Travel** (circle one)	**Project #**	**Supervisor Area #**	
	/ / 2011	Yes / No	See charts below	901	902

Each week record hours in this table, then total at the end of the week. Make corresponding notes as well.

Day of Week	Date	Production	Time									Notes
		# Comp. Interviews*	Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confer-ence	Other^	Daily Total	* Record QuestID for completed interviews ^ Explain 'Other' hrs.
Sun												
Mon												
Tues												
Wed												
Thur												
Fri												
Sat												
TOTALS												

Divide your total hours into the 1/3, 2/3 split. Record the splits in the first row below, then separate the charges between the two charts/codes so the splits and totals are close. Be sure the QuestIDs in the Notes correspond to the interviewing time recorded.

Day of Week	Date	MHSS **1/3: 0211838-212.006** 1/3 total: _____								EMHSS **2/3: 0212682-120.006** 2/3 total: _____							
		# Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total	# Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total
Sun																	
Mon																	
Tues																	
Wed																	
Thur																	
Fri																	
Sat																	
TOTALS																	

Notes:

Exhibit 7.1 ePTE Working Copy (continued)

Charges on ePTE:	Miscellaneous Expenses less than or equal to $10 Miles Driven Meal Expenses	Charges on Headway Expense Report:	Miscellaneous Expenses over $10 Lodging Costs, Auto Rental Telephone Charges

- Remember you must have approval from your Data Collection Manager before incurring any expenses.
- All Miscellaneous Expenses must be explained in the Notes. Expenses over $10 belong on the Headway Expense Report. (If you have multiple miscellaneous charges on one day that are <u>each</u> less than $10, but together total more than $10, use the ePTE and explain the charges in the Notes.)
- Divide total miles and total expenses using the 1/3, 2/3 split, and separate charges between the two charts.

Day of Week	Date	Expense							
		Miles Driven	Lodging#	Meal Expenses	Tele-phone#	Auto Rental#	Miscellaneous^	Incentive	Total Expense
Sun									
Mon									
Tues									
Wed									
Thur									
Fri									
Sat									
TOTALS									

Day of Week	Date	MHSS 1/3: 0211838-212.006				EMHSS 2/3: 0212682-120.006			
		Miles Driven	Meals	Misc.^	Total Expense	Miles Driven	Meals	Misc.^	Total Expense
Sun									
Mon									
Tues									
Wed									
Thur									
Fri									
Sat									
TOTALS									

Notes:

Advance Repayment

Do not use this box. You will not have an outstanding travel or incentive advance with Headway.

Notes

Type the following in the Notes section and press **[Enter]** after each item:

- QuestIDs of all completed interviews
- QuestIDs for any breakoff interviews that occurred and were not subsequently completed during the week. Enter BRK before the QuestID to indicate it is a breakoff case. You need not record breakoffs that result in a completed interview within the same week.
- Explanations of all time charges under "Other."
- Explanations of all Miscellaneous expenses $10.00 and under reported on ePTE.

When to Enter

You MUST record your ePTE information daily on the provided ePTE working copies. Record the information daily in the chart on the top half of the document as you work. Make notes to help you calculate your hours: record beginning and ending times as well as the stop and start times for any breaks taken. Make other helpful notes such as how long each interview took and any expenses paid. Then, when finished working, make your entries for each column in that day's row. Once you have finished working for the week, divide the data collection charges as explained in the next section, then transfer accurate data from your working copy to the ePTE system and submit it. Once a report is submitted, you cannot update it.

Remember, you must have your ePTE information entered into the ePTE system and submitted to RTI before **10:30 pm Eastern Time** on **Sunday**.

Dividing Charges

All CI data collection hours and expenses must be divided between two project numbers at the approximate ratio of 1/3 to MHSS and 2/3 to EMHSS. The ePTE Working Copy design helps in the division of charges. Follow these steps and consult the completed example provided in ***Exhibit 7.2***.

1. When finished working for a given week, total your hours in the chart on the top half of the ePTE Working Copy.
2. Divide the total hours into approximately 1/3 and 2/3, entering those values in the split charts at the bottom of the page, to the right of the MHSS/EMHSS project numbers. For example, if you work 12 hours, the split totals are 4 to MHSS and 8 to EMHSS.

   Total hours do not always divide cleanly into a 1/3 and 2/3 split, so round your divided totals to the closer quarter hour. The split does not have to be exact, but close. For example, if you work 14 hours, the split totals are 4.75 to MHSS and 9.25 to EMHSS.

3. Transfer the various amounts from your daily entries at the top to the split charts at the bottom so the totals reflect the 1/3 and 2/3 split. Consider these examples:
   - Divide a charge of 0.75 hours into 0.25 and 0.50.
   - If there are three separate charges for 0.25, record one on the first chart (MHSS) and the other two on the second chart (EMHSS).
   - Divide 1 hour into 0.25 and 0.75. However if you have three 1 hour charges, record 1 hour for MHSS and 2 hours for EMHSS.
   - Look at row totals for each day in the chart on the top to see if using daily totals makes it easier to divide, so that charges for one day are on one of the split charts and charges for another day are on the other chart. Use the smaller charges of 0.25 and 0.50 where needed to achieve the appropriate totals for each chart.
4. Balance your hours so the overall division works. If you have one charge that doesn't divide cleanly, look for another charge that also does not divide evenly, and balance these over the entire ePTE to achieve an approximate 1/3 and 2/3 split. Use the divided total hours recorded at the top of the split charts as your goal.

Carefully review the completed example in ***Exhibit 7.2***. This CI:

- Completed 3 interviews during the week while working 6.75 hours
- Divided 6.75 into the 1/3 and 2/3 splits, recording 2.25 for MHSS and 4.50 for EMHSS
- Looked along the daily row totals at the top of the chart, noting that for 8/26, the daily total was 1.75 (close to 2.25) while the total for 8/28 was 4.00, (close to 4.50)
- Recorded hours for 8/26 on the MHSS chart on the row for 8/26, and recorded hours for 8/28 on the EMHSS chart
- Used the remaining charges (conference, contact/locate) to balance the totals in each chart to end up with 2.25 hours for MHSS and 4.50 hours for EMHSS
- Added the appropriate QuestIDs and documentation for the 'Other' time in the Notes at the bottom.

Should you have any expenses, divide those entries into the 1/3 and 2/3 split as well using the charts on the reverse side of the ePTE Working Copy.

Remember to record the QuestIDs of any completed interviews in the Notes. Have the QuestIDs correspond (generally) to the interviewing hours charged.

Do your best to divide charges without spending too much time on the task. Your Data Collection Manager will provide any necessary feedback and is available to answer your questions.

Exhibit 7.2 Properly Completed ePTE

Clinical Interviewer WORKING COPY of the ePTE

Data Collection ePTE	Week Begin	Travel (circle one)	Project #	Supervisor Area #	
	08 / 22 / 2011	Yes / (No)	See charts below	901	(902)

Each week record hours in this table, then total at the end of the week. Make corresponding notes as well.

Day of Week	Date	Production: # Comp. Interviews*	Time: Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confer-ence	Other^	Daily Total
Sun	8/22										
Mon	8/23										
Tues	8/24					0.25					0.25
Wed	8/25					0.25					0.25
Thur	8/26	1				0.25	0.75	0.25		0.50	1.75
Fri	8/27					0.25			0.25		0.50
Sat	8/28	2				0.25	2.75	0.50		0.50	4.00
TOTALS		3				1.25	3.50	0.75	0.25	1.00	6.75

Notes
* Record QuestID for completed interviews
^ Explain 'Other' hrs.

9994321
Other: 8/26 + 8/28 audio, web entries, FedEx
9995432
9996543

Divide your total hours into the 1/3, 2/3 split. Record the splits in the first row below, then separate the charges between the two charts/codes so the splits and totals are close. Be sure the QuestIDs in the Notes correspond to the interviewing time recorded.

Day of Week	Date	MHSS 1/3: 0211838-212.006 (1/3 total: 2.25) # Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total	EMHSS 2/3: 0212682-120.006 (2/3 total: 4.50) # Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total
Sun	8/22																
Mon	8/23																
Tues	8/24			0.25					0.25								
Wed	8/25			0.25					0.25								
Thur	8/26	1		0.25	0.75	0.25		0.50	1.75								
Fri	8/27											0.25			0.25		0.50
Sat	8/28									2		0.25	2.75	0.50		0.50	4.00
TOTALS				0.75	0.75	0.25		0.50	2.25	2		0.50	2.75	0.50	0.25	0.50	4.50

Notes: 9994321 Other 8/26 audio, web, FedEx

9995432 9996543 other 8/28 audio, web, FedEx

Exhibit 7.2 Properly Completed ePTE (continued)

Research Triangle Institute PT&E Application - Entry Form (Version 3.0)

File Edit Tools Utilities Help

Electronic Production Time and Expense

000999 — RTI ID

Patience Supervisor — Supervisor Name &

321 — Area

08/22/2010 — Week Begin Date

0211838 — 212.006 — RTI Project Number

Travel Status ☐

PTE Type: Data Collection ◉ | Counting & Listing ○ | Supervisor ○

Did Not Work ☐

Employee: A. Professional

HEADWAY

Enter PT&E

Exit

Transmit ePT&E

RTI INTERNATIONAL

Research Triangle Institute PT&E Application - Detail Form

Data Collection PT&E

Project: 0211838-212.006 | RTI ID: 000999

Week Begin: 08-22-2010 | Travel: N | Supervisor: 321-Patience Supervisor

A		B	C								
Day of Week	Date	Production	Time								
		# of Completed Interviews*	Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confe-rence	Other*	Daily Total*
Sun	08-22										0.00
Mon	08-23										0.00
Tues	08-24					0.25					0.25
Wed	08-25					0.25					0.25
Thurs	08-26	1				0.25	0.75	0.25		0.50	1.75
Fri	08-27										0.00
Sat	08-28										0.00
Totals		1	0.00	0.00	0.00	0.75	0.75	0.25	0.00	0.50	2.25

B-1 Note

9994321

Other: 8/26 Audio, web entries, FedEx

* Click on labels denoted by "*" for more details

A		D							
Day of Week	Date	Expense							
		Miles Driven	Lodging	Meals Expense*	Tele-phone*	Auto Rental*	Misc. *	Incentive*	Total Expense
Sun	08-22								0.00
Mon	08-23								0.00
Tues	08-24								0.00
Wed	08-25								0.00
Thurs	08-26								0.00
Fri	08-27								0.00
Sat	08-28								0.00
Totals		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

$0.00 — Advance Repayment*

Save

Cancel

Print

Exhibit 7.2 Properly Completed ePTE (continued)

Research Triangle Institute PT&E Application - Entry Form (Version 3.0)

File Edit Tools Utilities Help

Electronic Production Time and Expense

000999 — RTI ID

Patience Supervisor — Supervisor Name &

321 — Area

08/22/2010 — Week Begin Date

0212682 — 120.006 — RTI Project Number

Travel Status ☐

PTE Type: Data Collection ◉ | Counting & Listing ○ | Supervisor ○

Did Not Work ☐

Employee: A. Professional

HEADWAY

Enter PT&E

Exit

Transmit ePT&E

RTI INTERNATIONAL

Research Triangle Institute PT&E Application - Detail Form

Data Collection PT&E

Project: 0212682-120.006 | RTI ID: 000999

Week Begin: 08-22-2010 | Travel: N | Supervisor: 321-Patience Supervisor

A		B	C									B-1
Day of Week	Date	Production	Time									Note
		# of Completed Interviews*	Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confe-rence	Other*	Daily Total*	9995432 9996543 Other: 8/28 audio, web entries, FedEx
Sun	08-22										0.00	
Mon	08-23										0.00	
Tues	08-24										0.00	
Wed	08-25										0.00	
Thurs	08-26										0.00	
Fri	08-27					**0.25**			**0.25**		0.50	
Sat	08-28	**2**				**0.25**	**2.75**	**0.50**		**0.50**	4.00	
Totals		2	0.00	0.00	0.00	0.50	2.75	0.50	0.25	0.50	4.50	

* Click on labels denoted by "*" for more details

A		D							
Day of Week	Date	Expense							
		Miles Driven	Lodging	Meals Expense*	Tele-phone*	Auto Rental*	Misc. *	Incentive*	Total Expense
Sun	08-22								0.00
Mon	08-23								0.00
Tues	08-24								0.00
Wed	08-25								0.00
Thurs	08-26								0.00
Fri	08-27								0.00
Sat	08-28								0.00
Totals		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

$0.00 — **Advance Repayment***

Save

Cancel

Print

When Finished Entering Data

If you want to exit the form without submitting (transmitting) it at this time, click the "Cancel" button on the Detail Form. A confirmation box appears asking if you want to save the changes you have made to the form. If you want to close without saving, click the "No" button. If you want to save the time sheet, click "Yes."

To save from within the Detail Form, click the "Save" button. The "Ready to Transmit" box appears:

If you have **not** completed the entire week's information for your ePTE, click the "No" button. A box then appears confirming for you that the ePTE has been saved, and can be edited and transmitted later. Press **[Enter]** to close this box.

If you are finished with this report and wish to send it to RTI, click "Yes." The following box appears:

YES: Answering "Yes" to this question certifies the charges you are submitting are accurate. This is like signing a paper document. Click "Yes" to sign and finalize your ePTE record.

NO: If you have any doubts about the information that you entered before you certify your information is accurate and correct, click "No." The next message box confirms for you that the ePTE has been saved, and can be edited and transmitted later. Click "OK" to close that message box.

Next you are asked if you want to enter another ePTE or transmit the data you have entered. If you are finished with the ePTE system, click "No" to exit. If you have additional tasks, click "Yes" and the system will return you to the Entry Form.

Transmitting Data to RTI

A finalized ePTE record marked as ready to transmit must be sent to RTI before 10:30 pm EST on Sunday. To transmit (send) your final ePTE information to RTI, follow these steps:

1. Be sure your computer is connected to the Internet.
2. Be sure your ePTE report is finished, and confirmed as "Ready to Transmit" through the "Save" process within the Detail Form (see previous page).
3. At the Entry Form, click "Transmit ePT&E."
4. The following text appears in a message box:

   > The system is going to transmit ePT&E records to RTI. Please press OK and wait until the program finishes before performing any other operations. If you do not intend to transmit at this time, press Cancel.

   If you do not wish to transmit, click "Cancel." To continue and transmit to RTI all ePTE records marked as ready to transmit, click "OK."
5. Wait patiently (i.e., don't try to multi-task) while the process takes place. When finished, the next message indicates the results.
6. **If successful**: The following text appears in a message box:

   > You have successfully transmitted your data. In this transmission you have transmitted x PT&E record(s) and y record(s) on address information.

   In the text, x and y represent the number of records sent to RTI. You may safely exit the system.
7. **If not successful**: A message box appears explaining the nature of the error and steps to take to correct the problem. Follow the instructions carefully. As needed, check with your Data Collection Manager.

During the transmission process, the record will be sent to RTI. Then, it will be processed, and forwarded to your Data Collection Manager for approval. Once the record is transmitted you can no longer edit the information. It becomes a read-only record which you can view but you cannot change (see ***Section 7.5.6***, item 4). If you need to make changes to a transmitted record, see ***Section 7.5.6***, item 5.

7.5.5 Edit Checks in the ePTE System

The program is sophisticated enough to provide some built in checks to prevent errors when filling out your ePTE. The system displays a message or warning box which identifies the problem and includes suggestions for resolving the problem. If you encounter a pop-up box, read it carefully. Press **[Enter]** to close the box and then continue.

The following are possible scenarios that cause an error pop-up message:

1. Begin work week date is in the future or is too far in the past.
2. Begin work week date is not a Sunday.
3. Letters and other characters can't be entered in a numeric field (such as hours worked).
4. You can't charge more than 20 hours in a day.

7.5.6 Advanced Features

The previous sections provide all the steps necessary to complete a regular ePTE report. This section includes additional system features and focuses on making changes to your ePTE data and your personal information.

1. **Editing Hours and Expenses on an untransmitted ePTE record:**

   You can open and edit any ePTE record in the system as long as the record has not been transmitted to RTI.

   - On the Entry Form, key in the exact information that you entered previously for that PT&E and click "Enter PT&E."

   You can use this method to update Hours or Expenses or other information in any un-transmitted records. Transmitting to RTI only sends those records that you specify as "Ready to Transmit" when answering "OK" to the certification question. If you choose not to transmit the record when you save it, by canceling or not certifying it, you must open it later using this method, and then save it with a transmission confirmation in order for RTI to receive the record before **10:30 pm Eastern Time** on **Sunday**.

2. **Deleting an ePTE Record that hasn't been transmitted:**

   To delete an un-transmitted ePTE record, key in the exact information on the Entry Form that you entered and saved previously. Once you have reached the Detail Form for the un-transmitted ePTE you wish to delete, click the "Delete" button at the top of the page. The system asks if you really want to delete this record. If you answer yes, the record is removed from the system. Please note that you can only delete records that are not yet transmitted to RTI.

3. **Locating Un-transmitted ePTE Records:**

   As explained above, you can locate and edit an un-transmitted record by providing the exact same information on the Entry Form as the record you want to locate. Sometimes, however, you may not remember the ePTE records that you have previously entered. The Entry Form provides a tool to help you. To find un-transmitted records:

   - Press [**Alt**] [**f**] (for File) then [**u**] for Search Un-transmitted ePTE.

     The PT&E Entries on File Form (shown on the next page) lists all ePTE entries in the system that have not been transmitted. Use the arrow keys to move up or down in the list until the record you wish to edit is highlighted in blue, then press [**Enter**]. Touchpad users can point to the needed record and double click to select it. This takes you back to the Entry Form where the fields are filled with the data from that particular record. Press [**Alt**] [**n**] to Enter the Detail Form (or click Enter) to make changes to the data associated with that ePTE record.

4. **Locating Transmitted ePTE Records:**

   Use the method from item 3 to view transmitted ePTE records. From the Entry Form, press [**Alt**] [**f**] (for File) and [**t**] for Search Transmitted PT&E to see a list of records that have already been transmitted to RTI. Selecting a record from this screen works the same way as it did with un-transmitted records. Use the arrow keys to highlight in blue the record you wish to view, then press [**Enter**], or double click the record. You can open these records for viewing purposes only. The system displays a pop-up box when you enter the Detail Form reminding you that the record has already been transmitted and asking you if you would like to open it for viewing only. All the fields in the Detail Form are grayed out or read-only, preventing you from making changes. Also, the only available option from the Detail Form is to Cancel, [**Alt**] [**c**].

5. **Correcting mistakes on an ePTE that has already been transmitted:**
   Please notify your Data Collection Manager as soon as possible if there is a problem with an ePTE you have submitted. You may be asked to contact Headway staff.

6. **Completing an ePTE for up to 2 weeks in which you do not work:**

   On the Entry Form, check the "Did Not Work" field. Then go to the Detail Form in order to save the record and make it "Ready to Transmit." If you check "Did Not Work" on the Entry Form, you are not able to enter any Hours or Expenses on the Detail Form.

7. **Completing a record for 3 weeks or more in which you do not work:**

   Do not submit an ePTE record if you will not be working for 3 weeks or more. Instead, your Data Collection Manager must send an e-mail message to Headway indicating the dates you will not be working as well as the reason you will not be working. This may occur in a number of situations, such as when you have finished your assignment and will not be working for the remainder of the quarter.

8. **Loading Default Values:**

   If you have made changes to the information on the Entry Form and now need to bring your default values back, press [**Alt**] [**t**] (for Tools) then [**l**] (for Load Default Values).

9. Clearing Default Values on the Entry Form:

Occasionally, you may want to clear out all default information displayed on the Entry Form to enter information for a new ePTE record. To do this, press [**Alt**] [**t**] (for Tools) then [**r**] (for Refresh).

10. Changing the Default Values on the Entry Form:

Occasionally, you may need to change the default values that pre-fill when you begin the ePTE program, including the project number.

- From the Entry Form, press [**Alt**] [**t**] (for Tools) and then [**v**] (for Update Default Values).

The System Default Value Setting Form appears so you can update your personal information. You can make corrections to all values except the RTI ID. If you make updates to your name, address, or phone number, a check mark appears in the "New Name/Address/Phone" field. After you make the updates, press [**Alt**] [**s**] (for Save) or click Save. A pop-up box asks if you would like to transmit this change. If so, answer "Yes" to this pop-up box. The next time you transmit to RTI this update information will be sent to RTI and Headway. Be sure to inform your Data Collection Manager of any name/address/phone changes as well.

11. Changing your RTI ID Number:

This ID number is the main key to processing your ePTEs (and getting paid!) and should only be changed under very unusual circumstances. To change your RTI ID number, you must acquire a password from your Data Collection Manager. Once you have the password, go to the Entry Form. Press [**Alt**] [**t**] (for Tools) and then [**i**] (for Change ID). The Change RTI ID Form appears:

Enter the updated information and the password, then press [**Alt**] [**o**] (for OK) or click "OK." If you have entered this form in error, cancel out by pressing [**Alt**] [**c**] (for Cancel) or clicking on "Cancel."

12. Compact your ePTE Data:

Occasionally, your Data Collection Manager or Tech Support may recommend that you compact your ePTE data to enhance the system performance and prevent system crashes. Compacting data will be scheduled by the person recommending the process and should not be done on your own. **When the system is compacting data, please wait until it finishes. Do not perform any other task or turn off your computer as this may cause serious problems**.

13. Accessing the Built-In Help Files:

From the Entry Form, press [**Alt**] [**h**] and then either [**b**] for Basic Information or [**q**] for the list of Frequently Asked Questions. You can scroll down the help window by using the **down arrow**. Once you have finished reading the help file, close it by pressing [**Alt**] [**F4**].

14. **Using the Touchpad:**

By design, this ePTE system works fine using the keys, although it has also been designed to be easy to use with a computer mouse. Your laptop has a built in Touchpad, which functions like a mouse.

7.6 Submitting Receipts with ePTEs

Most of your time and expense charges can be submitted using the ePTE. Since you should rarely have any expenses, you must check with your Data Collection Manager prior to incurring any expense. In the rare event you should have an expense over $10.00, prepare and submit a **Headway Expense Report** and include paper receipts to claim reimbursement. Bear in mind how each type/amount of expense is submitted and don't submit the same charge both ways. Every expense will be claimed on EITHER the ePTE OR the Headway Expense Report, but NOTHING will be reported on both of the forms. In other words, you cannot claim the same expense twice.

Follow these steps when using the Headway Expense Report (see ***Exhibit 7.3)*** to submit receipts:

Accurately record the area number, the project number, Data Collection Manager name and the corresponding week beginning date.

1. **Date Column** – Enter the dates beside each day of the week on the left side.
2. **Total Expenses Column** – Total your daily expenses by adding the amounts in each of the individual expense columns.
3. **Lodging Column** – For this study, you should not have any lodging expenses.
4. **Meals** – Meal expenses should be submitted on your ePTE, NOT on the Headway Expense Report.
5. **Telephone** – For this study, you should not have any telephone expenses since you will receive a project calling card. If needed, reimbursement for adding 3-way calling should be recorded under Miscellaneous Expenses.
6. **Auto Rental** – For this study, you should not have any auto rental expenses.
7. **Miscellaneous** – Document all miscellaneous expenses over $10.00 for the corresponding day of the week. Record the amount and explain the expense in the Notes section. **Receipts for all miscellaneous expenses over $10.00 must be submitted with this form**. Remember, expenses over $10.00 are not recorded on your ePTE.
8. **Notes** – Enter notes as required. For example, all miscellaneous expenses must be explained in the Notes section.
9. When finished, complete the certification area by entering your RTI ID number and the date prepared. Sign your name to "...certify that this statement is true, correct, and complete and that the data collected and submitted are truthful."
10. Keep the pink (bottom) copy for your records.
11. Staple all receipts behind the white copy of the form and mail the original and yellow copies to your Data Collection Manager.

Exhibit 7.3 Headway Expense Report

HEADWAY EXPENSE REPORT

RTI Project Number: _____

Field Supervisor Area ☐☐☐

Field Supervisor Name _____ Week Beginning Sunday ☐☐ Month ☐☐ Day ☐☐☐☐ Year

Day of Week	Date	Total Expenses	Lodging*	Meals	Telephone*	Auto Rental*	Miscellaneous§
Sunday							
Monday							
Tuesday							
Wednesday							
Thursday							
Friday							
Saturday							
	Totals:						

* All expenses for Lodging, Telephone and Auto Rental must be accompanied by original receipts.
§ All miscellaneous expenses must be explained in the Notes section. Attach receipts for expenses **over $10.**

Notes:

☐☐☐☐☐☐
ID Number

Name: _____ **(Please print)**

Address: _____ **(Street)**

City: _____ **State:** _____ **ZIP:** _____

Telephone: (_____) _____ - _____ ☐ **Please check if name or address change**

I certify that this statement is true, correct and complete and that the data collected and submitted are truthful.

Signature _____ **Date** _____

Approved by: _____ **Date** _____
Supervisor

Disposition: Send original and yellow copies to supervisor; retain pink copy for your records.

10/05

7.7 Authorization for Expenditures

Before spending any money on project-related business, consult with your Data Collection Manager to obtain authorization. This is to control budget expenditures and make sure the managers have input into all decisions affecting data collection work. Expenses made without prior authorization may be denied reimbursement. So you **must obtain approval prior to making any such expenditures**. You also must obtain and submit receipts for expenses over $10.00.

7.8 FedEx Mailing Procedures

You will need to send various materials or equipment via FedEx. For example, you will use FedEx to send completed, edited SCIDs to RTI within 48 hours of completing the interview. The procedures, while simple, are essential to ensure that your shipment is not lost, arrives promptly, and that the charge is billed properly.

When sending materials via FedEx, place materials in a package suitable for shipping. Your bulk supplies will include these packages, or your local FedEx office can provide you with boxes and envelopes free of charge if you need them. Follow the steps documented in ***Exhibit 7.4*** to properly complete the FedEx airbill. An example of the pre-printed FedEx airbill that will be provided for you to use for SCID shipments is shown in ***Exhibit 7.5***, while ***Exhibit 7.6*** contains an example pre-printed airbill for equipment shipments to RTI. It is especially important that you use one of the pre-printed FedEx airbills provided. If you need additional airbills, contact your Data Collection Manager to arrange for an additional supply to be sent to you.

7.8.1 Proper Handling of FedEx Packages

Project materials and equipment are vulnerable to loss or theft during drop off and delivery. Use the following steps to ensure proper delivery and handling of the package.

- Keep the top copy of the airbill, the "Sender's Copy," so you have a record of the tracking number at the top of the airbill. RTI staff need this number to track the package if it is not delivered on time.
- Peel off the adhesive backing and affix the label to the package, or place the completed FedEx airbill in one of the provided plastic pouches.

Once ready for shipment, consider the best method for giving the package to FedEx. Your options:

1. Arrange for the package to be picked up at your home by calling FedEx at 1-800-GOFEDEX. Since these packages contain confidential information, you MUST hand the package directly to an authorized FedEx employee. Do NOT leave the package unattended for pickup. Do NOT leave the package with a family member to give to FedEx.
2. Take the package to your local FedEx office. You must hand the package directly to an authorized FedEx employee. Do NOT ask any non-project person to deliver the package for you.
3. Leave the package in a secure FedEx drop box. Do NOT ask any non-project person to deliver the package for you.

While detailed, these FedEx procedures are designed to ensure that all project materials and equipment remain in "reliable hands" while in the field. Always ask if you have any questions.

Exhibit 7.4 Instructions for Completing a FedEx Airbill

Section #	Instructions	Additional Notes or Explanations
1	Enter the shipping date. Leave the Sender's account number blank. Enter your name, telephone number and address as the sender, if it is not pre-printed.	
2	Notice the preprinted project number, either 0211838.212.006 or 0212682.120.006. To divide FedEx charges into the 1/3 and 2/3 split, your stack of preprinted airbills is ordered with 1 MHSS airbill followed by 2 EMHSS airbills. Simply use the next airbill in the stack to properly divide the charges.	This preprinted number on the airbill must be marked correctly for RTI to properly process charges. For data security purposes, do NOT add a project name or acronym.
3	For **completed SCIDs**, use the airbill preprinted for RTI to . For **RTI equipment**, use the airbill preprinted for RTI Technical Support.	FedEx must have street addresses: no PO boxes.
4a	Verify that the box "FedEx Priority Overnight" is marked (unless otherwise instructed by your supervisor).	This must be marked correctly for RTI to receive the correct, discounted shipping rate.
4b	Leave blank.	
5	Mark the type of package used for the shipment.	"Other" should be marked for NSDUH equipment shipped in RTI packaging. If instructed by your supervisor to use FedEx Pak packing, make sure the appropriate package type is marked.
6	Verify that the "No" box is checked under the question about possible dangerous contents.	
7	Verify that the box "Third Party" is preprinted. Check also that the RTI FedEx account number ( ) is preprinted. Leave the weight/total declared value lines blank.	The standard FedEx coverage meets the project's needs. Additional insurance is not necessary.
8	Mark "Direct Signature" required.	A signature is required since a SCID package contains confidential information while a package to Tech Support contains valuable equipment.

Exhibit 7.5 Completed FedEx Airbill to RTI

FedEx Express US Airbill

Tracking Number 8583 1305 2551

Form ID No. 0215

Sender's Copy

PULL AND RETAIN THIS COPY BEFORE AFFIXING TO THE PACKAGE. NO POUCH NEEDED.

1 From Date 2/28/11 Sender's FedEx Account Number

Sender's Name Rachel Interviewer Phone (937) 618-5050

Company

Address 519 Valley Drive

City Oakwood State OH ZIP 45409

2 Your Internal Billing Reference 021838.212.006

3 To Recipient's Name Phone (

Company RTI INTERNATIONAL

Recipient's Address 1000 PARLIAMENT CT STE 100

Address

City DURHAM State NC ZIP 27703

/0800/1000/0034147661/6

4a Express Package Service

[X] FedEx Priority Overnight

[ ] FedEx Standard Overnight

[ ] FedEx First Overnight

[ ] FedEx 2Day

[ ] FedEx Express Saver

4b Express Freight Service

[ ] FedEx 1Day Freight

[ ] FedEx 2Day Freight

[ ] FedEx 3Day Freight

5 Packaging

[ ] FedEx Envelope

[ ] FedEx Pak

[ ] FedEx Box

[ ] FedEx Tube

[X] Other

6 Special Handling

[ ] SATURDAY Delivery

[ ] HOLD Weekday at FedEx Location

[ ] HOLD Saturday at FedEx Location

[X] No

[ ] Yes

[ ] Dry Ice

[ ] Cargo Aircraft Only

7 Payment Bill to:

[ ] Sender

[ ] Recipient

[X] Third Party

[ ] Credit Card

[ ] Cash/Check

Total Packages

Total Weight

Total Declared Value $ 100 .00

8 NEW Residential Delivery Signature Options

[ ] No Signature Required

[X] Direct Signature

[ ] Indirect Signature

519

Ship and track packages at fedex.com

Simplify your shipping. Manage your account. Access all the tools you need.

fedex.com 1800.GoFedEx 1800.463.3339

Exhibit 7.6 Completed FedEx Airbill to RTI Technical Support

FedEx Express US Airbill
FedEx Tracking Number 8583 1305 2551
1 From Date 2/28/11 Sender's FedEx Account Number
Sender's Name Rachel Interviewer Phone (937) 678-5050
Company
Address 519 Valley Drive
City Oakwood State OH ZIP 45409
2 Your Internal Billing Reference 0211838.212.006
3 To Recipient's Name TECHNICAL SUPPORT Phone (
Company RTI INTERNATIONAL
Recipient's Address 1000 PARLIAMENT CT STE 100
Address
City DURHAM State NC ZIP 27703
/0800/1000/0034147661/6
Ship and track packages at fedex.com
Simplify your shipping. Manage your account. Access all the tools you need.
fedex.com 1800.GoFedEx 1800.463.3339
Sender's Copy
0215
4a Express Package Service
FedEx Priority Overnight
FedEx 2Day
FedEx Standard Overnight
FedEx Express Saver
FedEx First Overnight
4b Express Freight Service
FedEx 1Day Freight
FedEx 2Day Freight
FedEx 3Day Freight
5 Packaging
FedEx Envelope
FedEx Pak
FedEx Box
FedEx Tube
Other
6 Special Handling
SATURDAY Delivery
HOLD Weekday at FedEx Location
HOLD Saturday at FedEx Location
No
Yes
Dry Ice
Cargo Aircraft Only
7 Payment Bill to:
Sender
Recipient
Third Party
Credit Card
Cash/Check
Total Packages
Total Weight
Total Declared Value
$ 100 .00
8 NEW Residential Delivery Signature Options
No Signature Required
Direct Signature
Indirect Signature
519
PULL AND RETAIN THIS COPY BEFORE AFFIXING TO THE PACKAGE. NO POUCH NEEDED.

This page intentionally left blank.

8. QUALITY CONTROL

8.1 Introduction

Your ability, preparedness, and willingness to properly perform your data collection tasks following all project protocols are the most important components of the overall quality control process. This chapter discusses the quality control procedures in place to help you do your best job.

8.2 Staff Training and Evaluation

During your training session, RTI project staff present you with the information necessary for you to properly perform your job as a clinical interviewer on this study. This includes detailed training on contacting interview respondents as well as a review of administrative procedures. Additionally, staff thoroughly cover the correct procedures for administering the SCID. ***Section 8.3*** describes procedures for certifying you in SCID administration. Additional information will be provided.

At the close of any in-person training session, you are asked to complete an evaluation form to assess the training program and materials, the trainers, and the training facilities. Following any formal group training conducted via telephone, please submit any feedback via e-mail to your Data Collection Manager. Your feedback on the effectiveness of the various interviewer training programs is an important part of letting us know whether or not the training program was thorough and effective. Your evaluation is also used to improve preparations for future training sessions.

8.3 CI Certification

To successfully graduate from training, each CI must complete and pass a rigorous certification interview to demonstrate his/her proficiency as a clinical interviewer. Following the completion of training, each CI will receive an assigned certification respondent, and complete a follow-up interview following all Mental Health Study procedures. Completed certification SCIDs received at RTI will be reviewed in detail by the Clinical Supervisors. Following this careful review, the CI will receive detailed feedback. If needed, the CI will be assigned one or perhaps two additional certification interviews.

MHSS cases may be assigned once a CI has been formally certified.

8.4 Supervision and Feedback during Data Collection

Feedback during this study focuses on helping you improve your clinical interviewing skills and on making sure the study runs smoothly. Feedback during data collection includes:

- discussions between you and your Clinical Supervisor concerning your performance while completing the SCID;
- discussions between you and your Data Collection Manager concerning logistical and administrative details, including any suggestions for ways to improve your performance of these tasks; and
- sharing with your Data Collection Manager your ideas and thoughts on ways to improve the study process (suggestions, whether through e-mail or during calls, are always welcome).

Please be receptive when given feedback from project personnel about your performance, as we are all striving to learn and improve.

8.5 Ongoing Quality Improvements

Throughout the year, typically before the start of each quarter of data collection, each CI must successfully complete a detailed inter-rater reliability exercise. Each CI will access an audio recording, listen to the recording, and complete a SCID based on the content of the recording. Details on how to access the file will be provided. After shipping the SCID to RTI, each CI's work will be reviewed with feedback provided individually by a Clinical Supervisor. These reliability exercises are important to maintain the standardization and integrity of the SCID administration process, thus maximizing the quality of the data collected.

Appendix A

NSDUH Respondent Materials

This page intentionally left blank.

To provide you with further background information on the NSDUH interview process, this appendix contains NSDUH materials and handouts used with respondents by field interviewers. These materials include:

- Question and Answer Brochure
- Summary of the Questionnaire
- Highlights from the 2009 NSDUH
- RTI/SAMHSA Fact Sheet
- Who Uses the Data?

For further information about NSDUH, SAMHSA, and RTI, visit the following websites:

http://nsduhweb.rti.org

http://www.oas.samhsa.gov/nsduh.htm

http://www.rti.org

Exhibit A.1 Question and Answer Brochure

National Survey on Drug Use and Health

Answering Your Important Questions

What Is the National Survey on Drug Use and Health (NSDUH)?

The National Survey on Drug Use and Health (NSDUH) is the Federal Government's primary source of national data on the use of alcohol, tobacco, and illicit substances. The survey also contains questions on health, illegal behaviors, and other topics associated with substance use. The study was initiated in 1971 and currently is conducted on an annual basis. This year approximately 70,000 individuals, 12 years old and older, will be randomly selected and asked to voluntarily participate.

The primary objectives of NSDUH are:

- to collect timely data on the magnitude and patterns of alcohol, tobacco, and illegal substance use and abuse;
- to assess the consequences of substance use and abuse; and
- to identify those groups at high risk for substance use and abuse.

Sponsored by

Substance Abuse and Mental Health Services Administration

U.S. Public Health Service

U.S. Department of Health and Human Services

Why Should I Participate?

- NSDUH is the primary source of national data on the use of alcohol, tobacco, and illicit substances. By volunteering in this study, you are helping us gather this important information that is needed to make accurate policy decisions.
- Individual residents of selected households, who are randomly chosen and agree to participate, are given a cash payment of $30 at the end of the interview.
- If selected to participate, you will represent over 4,500 other United States residents. Since our sample is selected based on scientific random sampling, no other household or person can be substituted.
- By participating in this study, you will be assisting with the formation of public policy.

Conducted by

Research Triangle Institute
3040 Cornwallis Road
Research Triangle Park, NC 27709

For more information on NSDUH or SAMHSA, contact:

NSDUH National Study Director
SAMHSA, Center for Behavioral Health Statistics and Quality
1 Choke Cherry Road
Room 7-1009
Rockville, MD 20857

For more information on NSDUH or RTI, contact:

NSDUH National Field Director
Research Triangle Institute
3040 Cornwallis Road
Research Triangle Park, NC 27709
1-800-848-4079

Internet Users: You may access more information about SAMHSA on the World Wide Web at:

http://www.samhsa.gov

Additional information about RTI is available at:

http://www.rti.org

Additional information about the National Survey on Drug Use and Health is available at:

http://nsduhweb.rti.org

V.1.11

Answering Your Important Questions about the **National Survey on Drug Use and Health**

What is the Substance Abuse and Mental Health Services Administration (SAMHSA)?

The Substance Abuse and Mental Health Services Administration (SAMHSA) is an agency of the U.S. Public Health Service in the U.S. Department of Health and Human Services (DHHS). SAMHSA was created in 1992 to provide leadership and a Federal focus for the Nation's mental health and substance abuse treatment and prevention programs. NSDUH is used to help facilitate this mission by monitoring the nature and extent of substance use in the United States, as well as the consequences of this use.

How Does the Government Conduct the Study?

Under a competitive bidding process, SAMHSA selects a survey research organization to administer NSDUH. Currently, Research Triangle Institute (RTI) is under contract to conduct NSDUH through 2011. RTI, which is located in Research Triangle Park, North Carolina, and closely associated with the University of North Carolina, Duke University, and North Carolina State University, is a large, experienced research organization that has conducted NSDUH since 1988.

How Was I Selected?

A scientific random sample of households is selected throughout the United States. Once selected, no other residence can be substituted for any reason. A professional RTI interviewer makes a personal visit to each household to ask several initial questions. One or possibly two residents of your household may be asked to voluntarily participate in the survey. If you are selected, no other person can be substituted. Since the survey is based on a random sample, you will represent over 4,500 other United States residents.

How Will the Data Be Used?

Government agencies, private organizations, individual researchers, and the public at large use the data for a number of purposes. For example, the U.S. Public Health Service and state health agencies use data from NSDUH to estimate the need for drug treatment facilities. Other federal, state, and local agencies use the information to support their drug use prevention programs and to monitor drug control strategies.

What if I Do Not Smoke, Drink or Use Illegal Drugs?

In order to know the percentage of people who do use these substances, we also have to know how many people do not. Therefore, the responses of people who do not use drugs are just as important as those of people who do. You do not need to know anything about drugs to answer the questions. In addition, we ask a number of health-related questions that are relevant for all people.

How Is the Study Administered?

NSDUH data are collected in the privacy of the participant's home. A professional RTI interviewer personally visits each selected household to administer the NSDUH questionnaire using a laptop computer. For some items, the interviewer reads questions and enters the responses into the computer; however, the participant privately enters most responses directly into the computer. The survey takes approximately 60 minutes to complete.

What Happens to My Information?

Each computerized interview data file—which is identified only by a code number—is electronically transmitted to RTI on the same day the interview is conducted. The answers then are combined with all other participants' answers, and are coded, totaled, and turned into statistics for analysis. As a quality-control measure, you may receive a telephone call or a letter from RTI to verify that the interviewer did complete the survey with you.

CONFIDENTIAL

Will My Answers Be Kept Confidential?

Both SAMHSA and RTI are committed to assuring complete confidentiality of responses. Our interest is only in the combination of all responses nationwide—not anyone's individual answers. Your full name is never recorded or associated with your answers. The information is only used for statistical purposes and cannot be used for any other purpose. Confidentiality of all answers to questions in this survey is assured under Federal law, the Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA). Any project staff or authorized data user who violates CIPSEA may be subject to a jail term of up to 5 years, a fine of up to $250,000, or both.

Exhibit A.2 Summary of the Questionnaire

2011 National Survey on Drug Use and Health
Summary of the Questionnaire

You have asked to know more about the ***National Survey on Drug Use and Health*** and the types of questions the interviewer will ask. Below is a summary of each section of the questionnaire for you to examine. Keep in mind that not everyone will see every question—the questions depend on the participant's own experiences. Furthermore, participants can always refuse to answer any questions during the interview.

Demographics
This section, administered by the interviewer, consists of questions about the participant such as his/her date of birth, race, educational background, and health status.

Sample Questions:
- ► What is the highest grade or year of school you have **completed**?
- ► Would you say your health in general is excellent, very good, good, fair, or poor?

Computer Practice Session
In this section, the interviewer shows the participant how to use the laptop computer and lets him/her practice using a short practice session.

Cigarettes, Alcohol and Illicit Drugs
For most of the rest of the interview, the participant answers questions by listening to the questions over the headphones and/or reading the questions on the computer screen, and then entering responses using the computer's keyboard. The participant answers these questions in private, although the interviewer is available to help with any problems using the computer. During this part of the interview, only the participant can hear and see the questions and see his/her responses.

Tobacco Products and Alcohol
These sections include questions about whether and how often participants have used cigarettes, chewing tobacco, snuff, cigars, pipe tobacco, and alcoholic beverages such as beer, wine, or liquor.

Sample Questions:
- ► Have you **ever** smoked part or all of a cigarette?
- ► How old were you the **first time** you used chewing tobacco?
- ► What is your **best estimate** of the number of days you drank alcohol during the past 30 days?

Illicit Drugs
The next sections ask about the participant's use or non-use of marijuana, heroin, cocaine, hallucinogens, and inhalants; and prescription pain killers, tranquilizers, stimulants, and sedatives when taken only for their effect. Questions about drug dependence and drug treatment are also included in these sections.

Sample Questions:
- ► Have you **ever**, even once, used marijuana or hashish?
- ► How much do people risk harming themselves physically and in other ways when they use **cocaine once a month**?

Exhibit A.2 Summary of the Questionnaire (continued)

Adult Social Environment and Mental Health
Participants aged 18 and older receive questions about their social experiences such as: how many times they have moved, their opinions about drug use, and in some cases, their experiences as a parent. Mental health questions cover such topics as depression and treatment for mental health problems.

Sample Questions:

- How do you feel about **adults** trying marijuana or hashish once or twice?
- During the past 12 months, was there any time when you **needed** mental health treatment or counseling for yourself but **didn't get it?**
- Have you ever in your life had a period of time lasting several days or longer when **most of the day** you felt **sad, empty, or depressed**?

Youth Experiences and Mental Health
Youth aged 12-17 participating in the survey are also asked questions about their social experiences such as: perceptions about the risks of using certain drugs; whether getting drugs is difficult or easy; feelings about school and peers; and involvement in clubs, sports, and other extracurricular activities. The mental health questions cover such topics as depression and treatment for mental health problems.

Sample Questions:

- **During the past 12 months**, in how many different kinds of school-based activities, such as team sports, cheerleading, choir, band, student government, or clubs, have you participated?
- Have you ever in your life had a period of time lasting several days or longer when **most of the day** you felt **sad, empty, or depressed**?
- During the past 12 months, did you receive treatment or counseling from a **private therapist**, **psychologist**, **psychiatrist**, **social worker**, **or counselor** for emotional or behavioral problems that were **not** caused by alcohol or drugs?

Health Care and Demographic Information
In this section, the laptop is handed back to the interviewer, who asks questions about education, health insurance, and family income information to help in analyzing the data. If necessary, a knowledgeable adult in the household may be asked to help participants answer some of these questions.

The answers to these questions increase the government's knowledge about health care, especially as it may relate to drug use or treatment. This information helps in planning health care services and finding ways to lower the costs of care.

Sample Questions:

- How many hours did you work **last week** at all jobs or businesses?
- **Before taxes and other deductions**, was the total **combined family** income during 2010 more or less than 20,000 dollars?
- Are you currently covered by private health insurance?

Please feel free to ask the interviewer if you have any other questions about the questionnaire.

Thank you for your cooperation and help!

Exhibit A.3 Highlights from the 2009 NSDUH

SELECTED HIGHLIGHTS from the *2009 National Survey on Drug Use and Health*

Tobacco Use

- An estimated 69.7 million Americans reported current use (during the past month) of a tobacco product in 2009, which is 27.7 percent of the population aged 12 and older. About 58.7 million (23.3 percent) smoked cigarettes.
- The graph to the right illustrates past month cigarette use among persons age 12 or older.

Past Month Cigarette Use among Persons Aged 12 or Older, by Age: 2009

Alcohol Use

Current Alcohol Use among Persons Aged12- 20, by Age: 2002-2009

+ Difference between this estimate and the 2009 estimate is statistically significant at the .05 level.

- Slightly more than half of all Americans age 12 or older, 51.9 percent or 130.6 million persons, were current drinkers in the 2009 survey, which is similar to the 129.0 million persons (51.6 percent) reported in 2008.
- Although consumption of alcoholic beverages is illegal for those under 21 years of age, 27.2 percent of this age group (10.4 million) were current drinkers in 2009. The graph on the left displays the current use of alcohol for 12–20 year olds from 2002 through 2009.

Illicit Drug Use

- An estimated 21.8 million Americans were current users of illicit drugs in 2009, meaning they used an illicit drug at least once during the 30 days prior to the interview. This estimate represents 8.7 percent of the population 12 years old or older.
- Marijuana is the most commonly used illicit drug, with an estimated 16.7 million current users, or 6.6 percent of the population 12 years old or older, an increase from the 2008 rate of 6.1 percent. Similar to 2008, an estimated 1.6 million persons were current users of cocaine, while 760,000 currently used Ecstasy, an increase from the 555,000 current Ecstasy users reported in 2008.

Results from the 2009 National Survey on Drug Use and Health: Summary of National Findings, DHHS/SAMHSA/CBHSQ, September 2010

Exhibit A.3 Highlights from the 2009 NSDUH (continued)

Mental Health

- In 2009, an estimated 14.8 million adults, or 6.5 percent of the population aged 18 or older, had at least one major depressive episode (MDE) in the past 12 months. Among adults, the percentage having MDE in the past year varied by age and gender, as shown in the graph below.

Major Depressive Episode in the Past Year among Adults Aged 18 or Older, by Age and Gender: 2009

- Persons with past year MDE were more likely than those without MDE to have used an illicit drug in the past year (29.5 vs. 13.5 percent).
- Similarly, substance dependence or abuse was more prevalent among persons with MDE than among those without MDE (22.4 vs. 8.2 percent), as shown in the graph to the right.

Substance Dependence or Abuse among Adults Aged 18 or Older, by Major Depressive Episode in the Past Year: 2009

Results from the 2009 National Survey on Drug Use and Health: Mental Health Findings, DHHS/SAMHSA/CBHSQ, 2010

Exhibit A.4 RTI/SAMHSA Fact Sheet

National Survey on Drug Use and Health

NSDUH

SAMHSA Fact Sheet

Substance Abuse and Mental Health Services Administration (SAMHSA)

About SAMHSA

The Substance Abuse and Mental Health Services Administration (SAMHSA), which funds the NSDUH, is an agency of the U.S. Public Health Service in the U.S. Department of Health and Human Services (DHHS). SAMHSA was created to focus attention, programs, and funding on improving the lives of people with or at risk for mental and substance abuse disorders.

SAMHSA is made up of four centers and supporting offices that engage in activities focusing on substance abuse treatment, mental health services, and substance abuse prevention. The Center for Behavioral Health Statistics and Quality (CBHSQ) is the focal point for the collection, analysis, and dissemination of national data on practices and issues related to substance abuse and mental disorders. This center is responsible for the NSDUH, the Drug Abuse Warning Network (DAWN), and the Drug and Alcohol Services Information System (DASIS), among other studies.

Selected Achievements

Recovery Month—Coordinated by SAMHSA's Center for Substance Abuse Treatment, the annual Recovery Month campaign highlights ways to create awareness that alcohol and drug use disorders can be managed effectively when an entire community supports those who suffer from these treatable diseases. Recovery Month promotes the societal benefits of treatment for alcohol and drug use disorders, recognizes the contributions of treatment providers, and highlights the message that recovery from alcohol and drug use disorders is possible. Annual NSDUH findings are presented during Recovery Month activities.
http://www.recoverymonth.gov

Access to Recovery—The Access to Recovery program, created and administered by SAMHSA, is an ongoing initiative that provides people seeking drug and alcohol treatment with vouchers to pay for a range of appropriate community-based services. Since the program began in 2004, more than $500 million in funding has been awarded to expand access to substance abuse treatment and recovery support services.
http://www.atr.samhsa.gov

Leading on Mental and Substance Use Disorders—SAMHSA submitted its landmark report on the long-standing barriers to appropriate treatment and support services for people having both mental and substance abuse disorders to Congress in 2002. Since then, SAMHSA has made significant gains in improving access to treatment for these individuals and continues to work toward further improvements in access and in the quality of services provided. These efforts include identifying and providing funding for effective programs in the prevention and treatment of both disorders, expanding SAMHSA's national registry of treatment programs, and establishing a national prevention and provider cross-training center to address this issue.

For more information about SAMHSA, contact:

NSDUH National Study Director
SAMHSA, Center for Behavioral Health Statistics and Quality
1 Choke Cherry Road
Room 7-1009
Rockville, MD 20857

http://nsduhweb.rti.org
http://www.samhsa.gov

Exhibit A.4 RTI/SAMHSA Fact Sheet (continued)

National Survey on Drug Use and Health

NSDUH

RTI Fact Sheet

Research Triangle Institute

About RTI

Research Triangle Institute (RTI) was created in 1958 through joint action by Duke University in Durham, the University of North Carolina at Chapel Hill, and North Carolina State University in Raleigh and has become an independently operating, nonprofit organization dedicated to conducting research that improves the human condition.

RTI is headquartered in the United States, with laboratory and office facilities in Research Triangle Park, North Carolina, and corporate offices and research facilities in California, Georgia, Illinois, Maryland, Massachusetts, Michigan and Washington, DC. RTI also maintains international offices in El Salvador, Indonesia, South Africa, Spain, Sweden, the United Arab Emirates, and the United Kingdom, as well as numerous project locations worldwide.

RTI offers innovative research and technical solutions to governments and businesses worldwide in the areas of surveys and statistics, health and pharmaceuticals, advanced technology, education and training, economic and social policy, international development, energy and the environment. Funds to support RTI's work come from competitively won contracts with and grants from clients in government, industry, and public service.

SAMHSA selected RTI to conduct the NSDUH through 2011. SAMHSA and RTI have worked together on the NSDUH since 1988.

RTI International is a trade name of Research Triangle Institute. Taxol®, a word coined by Dr. Monroe Wall of RTI, is a trademark of Bristol-Myers Squibb Company. Camptothecin™ is a trademark of RTI.

Selected Achievements

Taxol® and Camptothecin™—In the mid-1960s, two pioneering scientists at RTI produced two compounds instrumental in the fight against cancer—Camptothecin and Taxol—which have been responsible for saving the lives of hundreds of thousands of people afflicted with cancer.

Education Research—RTI is involved in education research and evaluation projects throughout the United States. In support of a contract awarded by the U.S. Department of Education to the Southern Regional Education Board, RTI has conducted work in the design, implementation, and evaluation of a comprehensive model for school reform in middle and secondary grades, with a focus on improving the quality of education programs and enhancing school and student performance.

Focus on Childhood Obesity—Recognizing the serious health and social consequences of obesity in children and adolescents, RTI completed two noteworthy efforts —with the University of North Carolina at Chapel Hill and through a grant from the Centers for Disease Control and Prevention—to explore public opinions about childhood obesity and lifestyle issues that may contribute to this problem. The results of these studies were utilized by policymakers to identify strategies and implement prevention programs where they are most needed.

For more information about RTI, contact:

RTI INTERNATIONAL

NSDUH National Field Director
Research Triangle Institute
3040 Cornwallis Road
Research Triangle Park, NC 27709

Phone: 1-800-848-4079
http://nsduhweb.rti.org
http://www.rti.org

Exhibit A.5 Who Uses the Data?

National Survey on Drug Use and Health
WHO USES THE DATA?

- The **White House Office of National Drug Control Policy** (ONDCP) uses the data to track progress toward goals in the National Drug Control Strategy.
- The **Substance Abuse and Mental Health Services Administration** (SAMHSA), a part of the U.S. Public Health Service, prepares statistical reports on substance use patterns and trends.
- **SAMHSA** uses the data to identify populations and geographic areas with particular substance abuse problems so that federal resources can be used efficiently for prevention and treatment programs.
- The **Partnership for a Drug-Free America** uses the data to design media advertising campaigns for the prevention of substance use and abuse.
- Based on the trends and patterns of substance use evident in the data, the **National Institute on Drug Abuse** (NIDA) develops research programs targeted toward populations and types of drug use problems where the need is greatest.
- **University-based researchers** use the data to conduct research on important substance use issues, such as the risk and protective factors associated with substance use, personal and societal consequences of substance use, and the impact of policy options for dealing with the substance abuse problem.
- **Substance abuse agencies at the state and local level** use the data to assess the potential need for treatment programs and to design programs that fit the needs of populations served.
- **State and local health departments** use the data to assess area substance use problems and to develop appropriate funding strategies and prevention measures.
- The **U.S. Department of Education** uses the data to inform drug use prevention and education programs and provide educational materials for teachers and administrators.
- The **U.S. Department of Transportation** uses the data on driving after alcohol and illicit drug use to develop prevention programs and materials on impaired driving.
- The **Office on Smoking and Health**, a part of the Centers for Disease Control and Prevention (CDC), uses the data to study trends and patterns in youth tobacco use and to develop strategies for reducing youth tobacco use.
- **Newspaper, television, and radio reporters** use the data in their stories on substance use and abuse.

For additional information on these organizations and their use of NSDUH data, please visit http://nsduhweb.rti.org and click on the "Who Uses NSDUH Data" link.

Appendix B

Instructions for Completing Paper PT&Es

This page intentionally left blank.

Instructions for Completing Paper PT&Es

As explained in ***Chapter 7***, all clinical interviewers must complete an electronic Production, Time, and Expense (ePTE) report weekly and submit it in order to be paid. In case of computer or system problems and you are not able to access the ePTE system, your Data Collection Manager may instruct you to prepare a paper version called an PT&E. Also, if a submitted ePTE is found to contain errors, a corrected paper report must be submitted.

1.1.2 What to Record on an PT&E Report

Step-by-step instructions for completing a paper PT&E report are provided in this section. Refer to ***Section 7.5.4*** for instructions on daily completion of the ePTE Working Copy (***Exhibit 7.1***) and splitting your hours/expenses between the different data collection project numbers.

An example of a properly completed paper PT&E Report is provided in ***Exhibit B.1***. Please refer to the instructions and the example as you are completing your paper PT&Es to be sure you are making all required entries correctly.

The information entered in the first 4 steps listed below defines an individual report.

1. Week Beginning Sunday – At the top of the form, enter the date (month, day, and year) of the Sunday on which the reporting period began. Each reporting period runs for seven days—Sunday through Saturday.

2. RTI Project Number – Enter the appropriate project number.

	Training/Certification	2011 Data Collection	Ratio
New CI	0212682.120.005	0211838.212.006	1/3
		0212682.120.006	2/3
Returning CI	0211838.212.005	0211838.212.006	1/3
		0212682.120.006	2/3

3. Did Not Work – If you do not work in a given week, you still must complete and submit a PTE report. Mark this box to indicate you did not work, completing the top of the form and the Certification Box in the lower middle of the form.

4. Section A (Date) – This section identifies the days by date over a one-week reporting period. Enter the corresponding date for each day shown.

5. Section B (Production) – Enter the total number of completed interviews (codes 70, 71, and 72). Add the entries in the column at the end of the week and record the total. You are also required to record the QuestID of each interview reported in Column B-1 in the "Notes" section, bottom-right. Make sure to keep up with your PT&E entries daily so you will know what you completed that day.

6. Section C (Time) – Entries should be made for each day you worked. In Column C-1, enter the total number of hours worked during the day as decimals, expressed in hours and portions of an hour to the nearest quarter hour. Show how your hours were spent on project work by distributing the time worked across the various columns as appropriate.

The "time" categories are defined as follows:

- Study/Training – Except for training, this column will not be used. However, if you are authorized to spend time doing additional study, record the time in this column.
- Travel – Use this column for any travel required for your work.
- Contacting/Locating – Record time spent contacting respondents, including time spent making appointments.
- Interviewing – Record time spent administering the questionnaire from start to finish including any initial conversations with the respondent during the interview call.
- Editing – Enter any time spent editing the SCID in this column. Do not include interviewing time here.
- Conference – Record time spent preparing for and in conference with your supervisors, compiling weekly reports or completing incident reports, and any calls with other RTI or research staff such as Technical Support.
- Other – Record time spent performing other allowable project activities such as entering information into the web as well as time spent transferring audio files and preparing FedEx packages. You must include a short description of any "Other" activity in the "Notes" section in the lower right corner of the report.

TOTAL THE ENTRIES IN ALL THESE COLUMNS AT THE END OF THE WEEK.

7. Section D (Expenses) – **Check with your Data Collection Manager before incurring any expenses**. With approval, enter the following for each day you work:

- Miles Driven – Record the total miles driven (rounded to the nearest mile) each day on project business (check to make sure that the corresponding travel time is shown in the Travel column). Mileage is reimbursed at the established government mileage reimbursement rate of $0.51 per mile. Note: This rate is set by the federal government and is subject to change.
- Total Expenses – Record the total amount of eligible expenses for each day.
- Lodging, Meals, Telephone, and Auto Rental – Do not use these columns unless instructed by your Data Collection Manager.
- Miscellaneous Expenses – Record all miscellaneous expenses. You must explain the expense in the Notes. Receipts must be submitted for expenses over $10.00—remember to get prior approval from your supervisor. Note that parking tickets, traffic tickets, and other fines are not reimbursable expenses.
- Incentives – You will not have any incentive expenses since respondents were paid for the telephone interview at the time of the initial interview.
- Please note the following:
  - Telephone – You will be using a provided calling card for your work and therefore should have no additional telephone charges to report. If needed, reimbursement for adding 3-way calling should be recorded under Miscellaneous Expenses.

8. Certification Section – Complete this section by printing your name, ID number, date, complete mailing address, and telephone number. Then sign and date the form, thereby certifying "that this statement is true, correct, and complete and that the data collected and submitted are truthful." Unsigned forms cannot be processed and will be returned, delaying payment to you.

9. Notes – The "Notes" section is for explaining *any* expenses and unusually high time charges to avoid delays in approving your PT&E. Also, record in the Notes section the following information:

- QuestIDs for all interviews completed during the week.
- QuestIDs for any breakoff interviews that occurred and were not subsequently completed during the week. Write BRK before the QuestID to indicate it is a breakoff case. You do not need to record breakoffs that result in a completed interview within the same week.
- Explanations of all time charges under "Other."
- Explanations of all Miscellaneous expenses (submit receipts for expenses over $10.00).

Maintain your PT&E report daily. Do not rely on your memory to complete it later. You will have photocopies of the ePTE that you can use as a working copy to keep track of your daily time and expenses, *including the QuestID numbers of completed and breakoff interviews*. Then, at the end of your work week, transfer the information to a "real" copy for submission to your supervisor. DO NOT SUBMIT THE WORKING COPY. IT CANNOT BE PROCESSED AND WILL DELAY YOUR PAYMENT.

Exhibit B.1 Properly Completed Paper PT&E

Clinical Interviewer WORKING COPY of the ePTE

Data Collection ePTE	Week Begin 08 / 22 / 2011	Travel (circle one) Yes / No	Project # See charts below	Supervisor Area # 901	902

Each week record hours in this table, then total at the end of the week. Make corresponding notes as well.

Day of Week	Date	Production # Comp. Interviews*	Time: Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confer-ence	Other^	Daily Total
Sun	8/22										
Mon	8/23										
Tues	8/24					0.25					0.25
Wed	8/25					0.25					0.25
Thur	8/26	1				0.25	0.75	0.25		0.50	1.75
Fri	8/27					0.25			0.25		0.50
Sat	8/28	2				0.25	2.75	0.50		0.50	4.00
TOTALS		3				1.25	3.50	0.75	0.25	1.00	6.75

Notes
* Record QuestID for completed interviews
^ Explain 'Other' hrs.

9994321
Other: 8/26 + 8/28 audio, web entries, FedEx
9995432
9996543

Divide your total hours into the 1/3, 2/3 split. Record the splits in the first row below, then separate the charges between the two charts/codes so the splits and totals are close. Be sure the QuestIDs in the Notes correspond to the interviewing time recorded.

Day of Week	Date	MHSS 1/3: 0211838-212.006 (1/3 total: 2.25) # Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total	EMHSS 2/3: 0212682-120.006 (2/3 total: 4.50) # Int.*	Travel	Contact Locate	Inter-viewing	Edit	Confer-ence	Other^	Daily Total
Sun	8/22																
Mon	8/23																
Tues	8/24			0.25					0.25								
Wed	8/25			0.25					0.25								
Thur	8/26	1		0.25	0.75	0.25		0.50	1.75								
Fri	8/27											0.25			0.25		0.50
Sat	8/28									2		0.25	2.75	0.50		0.50	4.00
TOTALS				0.75	0.75	0.25		0.50	2.25	2		0.50	2.75	0.50	0.25	0.50	4.50

Notes: 9994321 Other 8/26 audio, web, FedEx

9995432 9996543 other 8/28 audio, web, FedEx

Exhibit B.1 Properly Completed Paper PT&E (continued)

headway corporate resources

HEADWAY

under subcontract with

RTI International
Survey Research Division
PO Box 12194
Research Triangle Park, NC 27709

PRODUCTION, TIME, and EXPENSE REPORT (PT&E)

RTI Project Number: 0211838-212.006

Field Supervisor Area: 9 0 2

FS Name _____ Week Beginning Sunday 0 0 Month 2 2 Day 2 0 1 0 Year

Travel Status ☐ Did Not Work ☐

A		B	C								D							
		Production	Time								Expenses							
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Day of Week	Date	Number of Completed Interviews[1]	Total Hours Each Day[2]	Study/ Training	Travel	Contacting/ Locating	Interviewing	Editing	Conference	Other[3]	Miles Driven	Total Expenses	Lodging	Meals Actual Expense[4]	Tele-phone	Auto Rental	Misc.[5]	Incentives[6]
Sun	22																	
Mon	23																	
Tues	24		0.25			0.25												
Wed	25		0.25			0.25												
Thurs	26	1	1.75			0.25	0.75	0.25		0.50								
Fri	27																	
Sat	28																	
Totals		1	2.25			0.75	0.75	0.25		0.50								

B-1
Record ID #s of Interviews Completed During the Week
0994321

Notes
Other 8/26 audio, web, FedEx

0 0 0 9 9 9
(ID Number)

Name A. Professional Interviewer (Please print)
Address 1234 Main St (Street)
City Raleigh (Number) State NC ZIP 27529
Telephone (111) 222-3333 ☐ Please check if name or address change

I certify that this statement is true, correct and complete and that the data collected and submitted by me is truthful.

Signature A. Professional Interviewer Date 8/29/2010

Approved by _____ Supervisor _____ Date _____

Exhibit B.1 Properly Completed Paper PT&E (continued)

headway corporate resources

PRODUCTION, TIME, and EXPENSE REPORT (PT&E)

RTI Project Number: 0212682-120.006

under subcontract with

RTI International
Survey Research Division
PO Box 12194
Research Triangle Park, NC 27709

Field Supervisor Area: 902

FS Name _____ Week Beginning Sunday 08 (Month) 22 (Day) 2010 (Year)

Travel Status ☐ Did Not Work ☐

A		B	C										D							
		Production	Time										Expenses							
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Day of Week	Date	Number of Completed Interviews[1]	Total Hours Each Day[2]	Study/ Training	Travel	Contacting/ Locating	Interviewing	Editing	Conference	Other[3]	Miles Driven	Total Expenses	Lodging	Meals Actual Expense[4]	Tele-phone	Auto Rental	Misc.[5]	Incentives[6]
Sun	22																	
Mon	23																	
Tues	24																	
Wed	25																	
Thurs	26																	
Fri	27		0.50			0.25			0.25									
Sat	28	2	4.00			0.25	2.75	0.50		0.50								
Totals		2	4.50			0.50	2.75	0.50	0.25	0.50								

Report actual meal expenses with receipts and daily limits.
Total allowed meal expenses per day: $30.00
Incentive receipts must be attached

B-1	Notes
Record ID #s of Interviews Completed During the Week	other 8/28 audio, web, FedEx
99995432	
99965 43	

00099 9

Name A. Professional Interviewer (Please print)

Address 1234 Main ST (Street)

City Raleigh (Number) State NC ZIP 27529

Telephone (111) 222-3333 ☐ Please check if name or address change

I certify that statement is true, correct and complete and that the data collected and submitted are truthful.

Signature A. Professional Interviewer Date 8/29/2010

Approved by _____ Supervisor Date _____

Submitting a Completed Paper PT&E

Keep the gold copy of your paper PT&E for your records and FedEx the original, yellow, and pink copies (with any receipts stapled behind the while copy) to your Data Collection Manager no later than *Monday* of each week for review and signature. Once he/she receives your PT&E, he/she must review, approve, and sign it, and then send the forms to Headway/RTI for processing. All approved PT&Es received at Headway/RTI by 3 p.m. on Tuesday are processed, and checks or direct deposit vouchers are mailed by first-class mail to your home on Friday.

Please be aware that processing PT&Es takes time—sometimes as long as two or three weeks from the time you send the PT&E to your Data Collection Manager to the time you are paid, depending on holidays. Note that direct deposit is available to you, with funds available the Monday following Headway processing.

The following are examples of common errors found on paper PT&Es:

- Unsigned
- Failure to enter the project number (or, failure to enter the correct project number)
- Failure to enter a complete/correct ID Number
- Incorrect totals or no totals in columns and rows
- Breakdown of time not recorded in the proper column or totaled in a column other than the one in which daily entries are recorded.
- Time not rounded to quarter hours and not recorded in decimal form (two decimal points). Do not record time in minutes. Round to the nearest quarter hour:

Time Worked	Incorrect Entry	Correct Entry
1 hour and 15 minutes	1:15	1.25 hrs.
1 hour and 20 minutes	1:20	1.25 hrs.
1 hour and 30 minutes	1:30	1.50 hrs.
1 hour and 40 minutes	1:40	1.75 hrs.
1 and 3/4 hours	1:45	1.75 hrs.

Taking time to double check your work to catch any mistakes will save processing time and allow you to receive payment on time. Legible work helps as well—please write neatly in blue or black ink.

This page intentionally left blank.

Appendix C

The NSDUH Laptop

This page intentionally left blank.

C. THE NSDUH LAPTOP

C.1 Introduction

This chapter introduces the computer used for the Mental Health Study and serves as a reference whenever you have questions. It contains information about computer terminology and the different components of the Gateway laptop computer as well as tips on computer care and maintenance. If you cannot find an answer to a question, please call your Data Collection Manager.

C.2 Welcome to the Gateway E475

The Gateway E475 Notebook is a powerful laptop computer with many sophisticated functions and uses. RTI and SAMHSA believe in using the best equipment to conduct the NSDUH and have invested thousands of dollars in these laptops.

The laptop computer consists of many different parts including switches, lights, and ports for attaching optional equipment. While you will not use all of these components, take the time to become familiar with the ones you will use.

Your project issued computer bag contains all necessary pieces of equipment to conduct your work tasks. This chapter describes in detail all of the pieces of equipment used, including:

- the laptop unit
- a black power cord/AC adapter that comes in two parts
- headphones
- a telephone handset recording control (used between the handset and the telephone)
- wall jack recording control (used between the telephone and the wall jack of your telephone).

C.2.1 The Front

- **Power Indicator Light**. This blue light next to the battery indicator light is on when the computer is on and off when the computer is off. If the light is blinking, the computer is in 'stand-by' mode. Press the Power Button again to turn the computer on again.
- **Battery Status Indicator Light**. This light indicates the battery status and is located at the very front of the computer to the left of the latch. You can also see this light when the screen is closed.

  Blue—the computer is plugged into an outlet; battery is fully charged.

  Purple—the computer is plugged into an outlet; battery is charging.

  Blinking Red—the battery charge is very low.

  Solid Red—there is a battery malfunction. Contact Technical Support for assistance. If the problem prevents you from working inform your Data Collection Manager of the situation.

  Light is off—the computer is either turned off and not plugged in or turned on but not plugged in so they system is using the battery for power. Either way, the battery is not recharging.
- **Wireless Network Switch**. To access the Internet via a wireless network connection, slide the switch to the right to the 'on' position. If not using wireless, the switch should remain in the 'off' position to avoid causing problems.
- **Microphone and Headphone Jack**. These jacks are labeled to show you where to plug in headphones, speakers, or a microphone. To disable the laptop speakers, ***plug the headphones*** *into the* ***Headphone jack****, which is to the left of the speaker on the right side.*

  When connecting the recording control to the laptop, use the microphone jack, which is to the left of the headphone jack.

C.2.2 The Left Side

- USB Port. The USB port will not be used.
- Ventilation Fan. The laptop's ventilation fan helps cool the internal components. If the air vents are blocked, the laptop may become hot enough to harm your skin. Always place the laptop on a flat, sturdy surface, such as a table, and make sure the ventilation fan slots are never blocked as the laptop may overheat and cause permanent damage.
- Kensington Lock slot, Monitor Port, IEEE 1394 Port, PC Card Slot and Memory Card Reader. These features will not be used.

C.2.3 The Right Side

- USB Ports. Additional USB ports will also not be used.
- Smart Card Reader. Not used on NSDUH.
- CD Drive. The CD drive is located on the side of the computer.
- Modem Jack. This port, for connecting a telephone line, is used for a dial-up Internet connection.
- Ethernet Jack. Use this port for a cable or DSL high speed Internet connection.

C.2.4 The Back

- Power Connector. This socket is for the power cord/AC adapter. Plug the round connector for the AC adapter power cord into this socket.
- Battery. This is where the battery is located in the computer.
- S-Video Out Jack. Not used on NSDUH.

C.2.5 The Underside

Front of Gateway

- Battery. This is where the battery is located in the computer.
- Battery Latch. This latch, labeled with a small picture of a battery, releases the battery pack to replace the computer's battery (see ***Section C.5***).
- CD Drive Latch. ***Do not use this latch***. The CD drive is installed. Do not slide the latch over and remove the CD drive.
- Customer Care (Identification) Label. This label shows the respective serial number and other information for the laptop unit issued to you. There are several other labels (not shown above) on the underside of the computer. On the cover of the closed laptop, you will find an important blue and gray label indicating the laptop belongs to the Federal government. This label is also used for identifying purposes by RTI staff. **DO NOT remove any of these labels**.
- Hard Drive Bay. This is where your hard disk drive is stored. The hard disk is where all the data and computer software programs are stored. **DO NOT attempt to remove the hard disk drive from the bay** unless instructed to do so by Technical Support staff.
- Docking Port. Not used on NSDUH.
- Battery Lock. Used in conjunction with the battery latch to release the battery pack.

C.2.6 Additional Equipment

You will receive several additional items for use on the Mental Health Study, including headphones and two recording controls.

Connect the headphones to the laptop (refer to ***Section C.3.1***) to disable the laptop speakers. This is important for interviewing privacy.

Use one of the recording controls to record the interview. Refer to ***Chapter 3*** for connection and use instructions.

WALL JACK RECORDING CONTROL		
Phone Service:	**Analog**	
Control location:	Between wall jack and telephone	
TELEPHONE HANDSET RECORDING CONTROL		
Phone Service:	**Digital**	
Control location:	Between handset and telephone	

C.3 Opening the Computer

When the computer is opened, the screen and keyboard are displayed.

<u>To open the laptop</u>:

1. With the front of the computer facing you, find the silver latch on the front center of the laptop.
2. Slide the latch to the right and open the cover, displaying the screen.
3. Adjust the screen to a comfortable viewing angle.

C.3.1 The Front with the Display Open

- **Screen**. The Liquid Crystal Display (LCD) screen provides clear, bright color with minimal glare.
- **Display Hinges**. These hinges allow you to adjust the angle of the screen for comfortable viewing. To avoid damaging the screen, do not force it back too far.
- **Power Button**. This is the silver button found at the top right of the open laptop above the keyboard. The button glows blue when the computer is on.
- **Microphone**. This will not be used on NSDUH.
- **Touchpad**. This dark gray square below the keyboard acts as a "mouse." ***Section C.3.4*** contains details about Touchpad use.
- **Touchpad Buttons**. The left and right buttons below the Touchpad act like "mouse" buttons. Again, ***Section C.3.4*** contains details about Touchpad use.
- **Keyboard**. The keyboard allows you to enter information into the computer. Do not alter the keyboard or surrounding area by adding any labels or stickers or 'cheat sheets' other than the approved and standardized function key labels provided. See ***Section C.3.3***.
- **Status Indicator Lights**. The series of indicator lights inform you of the computer's status. For more information about the function of each light, see ***Section C.3.2***.

C.3.2 Status Indicator Lights

The status indicator lights are a row of five lights above the keyboard and visible below the screen when the laptop is open. It looks like this:

- **Wireless Network Light**. This light is blue when the wireless network switch is in the 'on' position.
- **Lock Key Lights**.

  Caps Lock — This light shows the [Caps Lock] key is on, locking the alphabet keys in uppercase letters. Press the [Caps Lock] key to turn the light off and return to typing in lowercase.

  Number Lock will not be used.

- **Module Light**. This light is blue when the CD Drive Module is in use. You will not be using this module during interviewing.
- **Hard Drive Light**. This light flashes when the hard drive is in use.

C.3.3 The Keyboard

The computer keyboard, used to enter information, contains three types of keys: **typewriter keys, function keys**, and **special keys**.

Typewriter keys type numbers, letters, punctuation and some common symbols (e.g., $, %, &). Please note the letter "l" and the number "1" are not interchangeable, neither are the letter "O" and the number "0" so be careful when typing.

Function keys on the top row, labeled F1 through F12, instruct the computer to carry out special commands.

The **special keys** are all the remaining keys.

- *ENTER KEY*. The most important special key is the [ENTER] key used to indicate that entry of a certain item is completed.
- *SHIFT KEY*. There are two shift keys on the keyboard, on the left and right of the second row from the bottom. They are marked with an up arrow and the word Shift. There is no difference between the left- and right-shift keys. The shift keys can be used in the same ways as the shift keys on a typewriter. If Shift is pressed with [a], an "A" is displayed on the screen, and pressing Shift and [1] simultaneously, displays a "!" on the screen.
- *ALT KEY*. The [Alt] key is located at the bottom corner of the keyboard on either side of the Space bar. It is also used in combination with other lettered keys on the keyboard to activate certain functions.
- *TAB KEY*. The [Tab] key is located on the left side of the keyboard on the third row from the top. It is used to move around the screen to choose an option.
- *ARROW KEYS*. On the lower right corner of the keyboard there are four arrow keys that can help you move to a particular area on the screen.

C.3.4 Touchpad Use

The Touchpad and Touchpad buttons shown on the previous page work like a computer mouse.

MOVE: The dark gray rectangle is a touch sensitive area. Placing your finger gently upon this rectangle and moving your finger around causes the cursor—the arrow on the screen—to move in the same direction as your finger. If you need to move the cursor a large distance, you may run out of room on the Touchpad. Simply lift your finger and move it to a different spot on the Touchpad then continue to move it.

POINT: Use your finger on the Touchpad to move the cursor to point to one of the words or boxes or commands shown on the CAI Manager screen.

PRESS: With the cursor pointing to a specific word/box/command on the CAI Manager screen, you must tell the computer you want to execute that step. Similar to pressing the Enter key, press the left button below the Touchpad to execute.

Computer linguists simplify the move-point-click steps by using just one word—click. Additionally, you can 'double click,' which means to press the left button twice in succession.

C.4 Powering the Computer

This section contains the necessary information to safely use electricity either from a wall outlet or from a battery to power your laptop. **Whenever possible**, the computer should be plugged into an outlet and should not be using battery power. Plugging in the laptop increases the screen brightness, enabling you to see the screen clearer and conserves battery power.

C.4.1 Using Electricity/AC Adapter

To plug the unit into an outlet, use the AC adapter power cord. It comes in two parts: (1) a power cord with an electric plug at one end and a connector at the other; and (2) a cord with the adapter "block."

To connect using the AC adapter power cord:

1.) Connect the power cord with the electric plug at one end to the AC power block. Use only the power cords that come with the laptop.

2.) Plug the complete AC adapter cord into the computer's circular power connector on the back of the laptop. When looking at the back, the connector is on the left side.

3.) Connect the other end of the power cord to a live wall outlet. Note: You should always plug the round end of the cord into the laptop first, and then plug the pronged end into the wall.

Warning! Do not disassemble the AC power block. It contains dangerous voltages that can cause serious personal injury or death. Contact your Data Collection Manager about returning defective adapters.

C.4.2 Using Battery Power

The battery that came with your computer arrived partially charged and ready to use; it is already installed in the laptop. To charge it completely, connect the computer to an outlet (see ***Section C.4.1***). The battery charges while your laptop is on or off if the cords are properly connected. When the computer is on and plugged into the wall, the battery recharges but at a slower rate than when the computer is off and plugged into an electrical outlet. Therefore, always recharge the battery overnight by turning off the computer, plugging in the power cord, and connecting to a wall outlet.

Although it is best to conduct interviews with the laptop connected to electrical power, the laptop generally can run between 2 to 3 hours on a fully-charged battery. Keep in mind that if not plugged in the battery gradually loses its charge even when you are not using the computer. The battery is fully charged when the battery indicator light is blue. To check the battery status, move the cursor over the plug/battery icon in the system tray in the lower right corner of the screen. When not plugged in, the remaining battery power percentage displays.

To recharge the battery:

1.) Plug the AC adapter power cord into the computer and an outlet, as described in ***Section C.4.1***.

2.) The battery will begin recharging immediately even if the laptop is running. Check to make sure the battery light is blue or purple, which indicates that the outlet is providing electricity to the computer.

Warning signs when the battery level is low:

- ***System Warning***. The computer emits three short "ding" sounds.

  If you hear this sound, your battery is getting very low. You must switch to electrical power as soon as possible. Plug the laptop into an electrical outlet, and proceed.

  Do not press the power button to turn off the laptop first—if you do, you may lose information. But it is not necessary nor advisable to shut down the laptop prior to plugging it into electrical power.

  When you hear the ding sounds, you must plug in your laptop immediately. The laptop will emit three more "ding" sounds as the battery is nearly empty and then *will* shut itself off automatically. If an outlet is not available or you cannot plug the laptop in for some other reason, stop and save what you are doing, and shut down the laptop.

C.4.3 Turning the Computer On

Once you have supplied power to your laptop by either using the battery or plugging into an outlet, starting the computer is simple.

To turn it on:

1.) Push the latch on the front of the laptop and open the display screen.

2.) Press the silver power button on the right of the computer, above the keyboard. [*Note: When you first press the power button, the computer's internal fan begins running. This noticeable noise from the fan may be startling at first, but is a normal function of the computer.*]

3.) Wait for the computer to boot up. You will know it is ready when the security password screen appears.

C.4.4 Turning the Computer Off

When turning off the computer, follow the steps shown below to avoid losing information or damaging the computer's hard disk drive. Do not use the power button to turn off the computer unless instructed to do so by Technical Support staff.

<u>To turn the computer off:</u>

1.) Simultaneously press the Ctrl/Alt/Delete keys, then choose Shut Down.

OR

2.) Click the Start button, select Shut Down, then click OK

Wait a moment or two for the computer screen to go blank before closing the display.

<u>IMPORTANT</u>: Do <u>not</u> push the power button at this point or you will turn the laptop on again.

C.5 Reseating/ Replacing the Battery Pack

If your battery is not charging correctly (either it is not charging at all or not holding a charge long), you can try to reseat the battery pack by taking it out and putting it right back in. In rare instances, a problem with the battery pack may require that it be replaced. Check with Technical Support staff for assistance.

To reseat or replace the battery pack:

1.) Turn the computer off.

2.) Close the computer's screen until it clicks shut and turn the computer over.

3.) Unplug the computer if it is plugged into an outlet.

4.) Slide the battery lock to the unlocked position, then slide the battery release latch. Slide the battery carefully out of the laptop.

5.) Place the battery pack (either the current one or a new one) into the battery bay until it "snaps" into place. Slide the battery lock to the locked position. Replacement battery packs are supplied by Technical Support.

6.) Turn the computer right side up, plug the computer back into electrical power, and then open the cover.

7.) Press the power button to turn the computer on.

C.6 Care of the Laptop

The laptop is an expensive piece of equipment that needs to be properly cared for so that it operates correctly. Take time now to read and follow these instructions to save time and effort later.

C.6.1 Exposure to the Elements

- **Keep the computer away from sources of extreme heat or cold**. Extreme temperatures can damage the computer. Do not place or use the computer close to a heater or air conditioner. **Store the computer in temperatures between 40 F and 95 F**. If the computer has been exposed to extreme temperatures, allow enough time for it to return to room temperature before use.
- **Avoid areas with high concentrations of dust and dirt**. These can clog the internal mechanisms of the computer and the keyboard.
- **Use the computer where it will have adequate ventilation**. Like many household appliances, the computer requires air circulation to keep it from overheating during use. While the circulation required is minimal, be sure you have not covered the computer's air vents during use.
- **Do not eat, drink, or smoke when using the computer**. Exposing the computer to food particles, liquids, or smoke may damage the computer permanently.

C.6.2 Screen, Keyboard, and Body Care

- **Do not set anything on the computer keyboard**. Accidentally closing the screen with something on the keyboard (such as a pencil), damages the computer screen.
- **Do not place heavy objects on the computer when it is closed**.
- When moving or transporting the computer, take special care not to bump it into objects. The more jolts the computer receives, the sooner it is likely to malfunction. **A single shock can potentially damage the computer permanently**.
- Always use the carrying case when taking the computer anywhere. The case is not only convenient for carrying the computer; it also helps protect it from damage.
- **Never pick up the computer by the screen**. The computer screen is the weakest part of the machine. Picking the computer up by the screen when it is open risks serious damage.
- **Never spray any liquid directly onto the computer**. To clean your computer, use a clean, damp cloth. You should apply a small amount of water or a gentle cleaning solution to the cloth and not spray it directly onto the computer. If you spill liquids on the computer, wipe off the liquids and allow the computer to dry thoroughly before using it again. Failure to do so will likely result in damage to the machine.
- **Use only water to clean the screen**. Other cleaning solutions can damage the screen despite advertisers' claims.

C.7 Cables, Ports, and Card Slots on the Laptop

- **Do not use frayed or otherwise damaged cables**. When connecting or disconnecting a cable, hold the cable only by its connector—the plug—not by the cord. If you notice a damaged cable, promptly contact Tech Support for a replacement.
- **Never force a connector into a port if the connector and port do not join easily**. Make sure the connector matches the port and that it is properly aligned before you attach it, sliding it straight into and out of the port. Attaching/detaching a cable at an angle will likely cause damage. Also take care not to mistakenly insert any other objects that do not belong (such as pens, pencils, etc) into any ports or openings.
- **When transporting your laptop, always disconnect any cables or cords**.

C.8 Traveling with your Computer

C.8.1 Air Travel

- When traveling by air with the laptop, **you can safely put your equipment through airport security devices that do not harm laptop computers**. If the security device is not marked that it is safe for computers, ask. If airport security personnel will not guarantee the safety of your laptop, hand it to the Security Guard, who will follow their standard protocol for checking computer equipment.
- **Never allow anyone to check your computer unless they have presented official airport security identification**. Beware of imposters who might try to trick unwary travelers into allowing them access to or control of their computers or other valuable items. Also, beware of two-person or larger teams working to get between you and your computers. Keep your laptop close to your body at all times. Do not set it down or let someone take it from you without proper identification. Another suggestion is to put your personal carry-on items through the scanning machine first, sending your laptop through last as you walk through the detection equipment. Doing so helps keep your computer equipment close to you.
- **Never check your laptop in your luggage**. Expensive electronic equipment should **always** be taken on the airplane as carry-on luggage. If it is checked with the rest of your luggage, the chances of it being stolen or damaged are *tremendous*.

C.8.2 Automobiles

- Never store your computer in the trunk of your car for any length of time. Doing so might expose it to extreme temperatures that can cause serious harm or permanent damage.
- Never leave your computer alone in the car where it is visible at all. This is an invitation to a thief to break into your car and steal it. Remember, you are responsible for returning your equipment to RTI when your data collection duties are over.

C.9 Your Responsibility

By understanding and following the care instructions given, you can take proper care of the equipment entrusted to you. The equipment issued to you is **your** responsibility. It is expected you will take the necessary precautions to ensure its safety and wellbeing. Additionally, the equipment is assigned to you only for your use as an interviewer on this project.

When you receive your equipment, you complete and sign an Equipment Assignment and Receipt Form, referred to as an EARF (see ***Exhibit C.1***). The notices at the top on the form contain important information. You must read each of these notes carefully and sign indicating you have read and understand the information.

- The first note acknowledges your receipt of the equipment and concerns use of and alterations to the equipment. You agree not to install, or attempt to install, any other software or programs on the equipment unless you are specifically told to do so by Technical Support staff. Also, you are not to alter or attempt to alter the existing programs. Additionally, if requested by project management or Headway, you are to return project equipment immediately to the appropriate person(s) designated by your Data Collection Manager.
- The second important notice concerns your responsibility for caring for the equipment. You are expected and legally responsible for taking appropriate care of the equipment, following the care guidelines listed in this appendix and taking steps to ensure the safety of your equipment. What does that mean? Basically, be careful! Don't repeatedly jam cords into connectors causing damage. Don't leave your computer in your car in plain sight. It is important that you follow the guidelines and be sensible when working with your equipment.

  While it is understood that an occasional accident or event out of your control may occur, Headway has specific consequences for theft, loss, or damage beyond repair, including disciplinary steps and required repayment of the approximate value of the equipment for repeated problems. These consequences are documented in the Headway Equipment Policy included with your hire packet.
- The next two notices acknowledge your receipt of the Headway Equipment Policy, confirm you have reviewed the policy, and that by signing you agree to the terms stated in the EARF and to adhere to the Headway Equipment Policy at all times.
- The last notice at the top of the EARF emphasizes the importance of keeping track of your government-issued equipment and lists the federal penalties for loss.

When you receive equipment and sign the EARF, you assume full and legal responsibility for taking reasonable and appropriate steps to safeguard this equipment against damage, loss, or theft. Should you experience loss or theft of your equipment or damage requiring repair, Headway will implement the disciplinary steps as documented in the Headway Equipment Policy, beginning with a letter of documentation.

Also, you are only to use the equipment provided to you for your work on the Mental Health Study. Do not add other software or games or even play music in the laptop CD drive. Use in this way is likely to cause problems with the specially designed computer programs. Should you attempt to alter the programs or load other software, you will be subject to Headway disciplinary procedures, beginning at the

Written Advisory (warning) stage (see the Headway Equipment Policy for more details).

SAMHSA has provided you with excellent equipment for your use on NSDUH. You are responsible for the proper handling and care of this equipment. Swapping of equipment with other staff is not allowed under any circumstances.

Exhibit C.1 Equipment Assignment and Receipt Form (EARF)

National Survey on Drug Use and Health

Project 0211838; Contract No. HHSS283200800004C

EQUIPMENT ASSIGNMENT AND RECEIPT FORM - HEADWAY

Directions: Read this Equipment Agreement carefully. Verify that each piece of equipment issued to you is properly identified on this form. Sign and date within each box for items checked out.

I acknowledge receipt of the following equipment which is in working order and new/like new condition. I understand that all equipment provided to me is the property of the United States Government and is to be used only for NSDUH project business purposes. I agree that while this equipment is in my possession, I will not in any way alter the software that was installed by RTI, unless specifically instructed to do so by RTI staff. I agree not to install any software on this equipment that is not supplied by RTI. Additionally, upon request by project management or Headway, I will return all equipment immediately to the appropriate person(s) as designated by my Field Supervisor. I understand that failure to return project-related equipment will result in disciplinary action, legal action, and/or payroll deductions.

Furthermore, I understand that I am fully and legally responsible for taking reasonable and appropriate steps to safeguard this equipment against damage, loss, or theft. In the event that equipment becomes damaged, lost or stolen, I will notify project management and Headway immediately, and in the event of theft, I will cooperate with any police investigation.

I acknowledge and confirm that I have received a copy of Headway's Equipment policy. I have reviewed this policy and directed any questions pertaining to the policy to my immediate supervisor or a member of the onsite team at RTI. I agree to adhere to the policy at all times. I further understand that if this agreement is not upheld in any manner that disciplinary actions will be taken immediately.

This agreement is binding and shall supersede all other agreements whether in writing or verbal and shall be governing terms and conditions where a conflict may exist between the terms and conditions of this agreement and the terms and conditions of other agreements or forms. As certified by my signature below, I hereby affirm that I have read, understand, and agree to the above terms and conditions, as well as to the Headway Equipment Policy.

NOTICE: The computer and the other equipment which will be provided to you for your use on this project are the property of the United States Government and must be returned immediately upon the completion of the assignment or upon termination of your employment for any reason. Failure to return the equipment is conversion of United States Government property and is a federal crime under the provisions of Article II, Section 641 of the United States code which provides that any person who embezzles, steals, purloins, or knowingly converts to his use or the use of another, or without authority sells, conveys, or disposes of anything of value of the United States shall be fined in an amount up to $10,000 or imprisoned not more than 10 years or both. Any person failing to return equipment will be reported to the Federal Bureau of Investigation and local law enforcement authorities.

Signature _____ **Date** _____

CHECKED OUT TO: (Print Name) _____

ID # _____

Place Label Here

iPAQ Kit (check all that apply)

- ☐ iPAQ (**2011 Value = $192**) Govt. ID # _____
- ☐ AC Adapter
- ☐ Car Charger
- ☐ Rechargeable Battery Pack
- ☐ Storage Card
- ☐ Spare Pens (2)
- ☐ iPAQ Canvas Case
- ☐ Mini-USB Cables (2)

Signature _____

Date _____

Laptop Kit (check all that apply)

- ☐ Laptop (**2011 Value = $620**) Govt. ID # _____
- ☐ AC Adapter
- ☐ Headphones
- ☐ Black Extension Cord
- ☐ CD Drive
- ☐ Phone Cords (2)
- ☐ Phone Cord Coupler
- ☐ Laptop Case

Signature _____

Date _____

Other Equipment (check all that apply)

- ☐ Printer (2011 Value = $153) Govt. ID # _____
- ☐ Monitor (2011 Value = $125) Govt. ID # _____
- ☐ Projector (2011 Value = $461) Govt. ID # _____
- ☐ Other _____

Signature _____

Date _____

Prepared/Assigned By: TSG Staff: _____ Date _____

RTI – NSDUH Technical Support Group, 1000 Parliament Court, Suite 100, Durham, NC 27703

Disposition: Original to RTI project files, yellow to Headway files, retain pink copy for your records. 08/10

Appendix D

SCID Specifications

Acknowledgements

Material presented in this appendix has been adapted from the User's Guide for the Structured Clinical Interview for DSM-IV-TR Axis I Disorders (Research Version) by Michael B. First, M.D., Miriam Gibbon, M.S.W., Robert L. Spitzer, M.D., and Janet B. W. Williams, D.S.W.

I. Overview

Before systematically inquiring about the presence or absence of particular DSM-IV criterion items, the SCID begins with an open-ended Overview of the present illness and past episodes of psychopathology. The Overview provides an opportunity for the respondent to describe his/her current situation in his/her own words. It also provides the interviewer the opportunity to collect certain types of information that may not be covered in the course of assessing specific diagnoses (e.g., prior treatment, social and occupational functioning, context of the development of the psychopathology). By the end of the interview, the interviewer should have gathered enough information to formulate a tentative differential diagnosis.

The Overview does not assume a chief complaint (current problem); it inquires about current and past emotional problems and any treatment that the respondent may have ever had. If the respondent has never had any kind of psychiatric treatment, two kinds of questions are particularly important for locating periods of possible lifetime mental disorders: 1) Has there ever been a time when the respondent was unable to go to work or school? And 2) When, in his/her life was the respondent most upset?

When asking about a history of past treatment and it becomes clear that the respondent has had a particularly complicated history, it may be useful to turn to the Life Chart, located at the end of the Overview. This chart provides a framework for recording past treatment history in a chronological fashion.

In the Overview, respondents are asked about all past treatments, including medications. The interviewer should be sure to question a respondent about any medications that were prescribed that do not seem appropriate for the condition described. This often gives a clue to problems that the respondent has not mentioned. For example, a respondent who describes only chronic depression, but was treated with lithium in the past, may describe a possible manic episode when asked why lithium was prescribed. (Of course, a prescribed medication should NOT be used to justify a diagnosis without documentation that the disorder actually met criteria, since medication is sometimes prescribed inappropriately.) Be sure to get a list of ALL psychotropic medications taken in the past 12 months, not just currently. Obtain the dosage, frequency, and any changes in dosage of all medications. Ask what symptoms or disease a medication is prescribed to treat. Many medications are prescribed for very different disorders

(e.g., Depakote for both seizures and mania; Paxil for both the pain of fibromyalgia and depression); dosages often differ depending on the symptoms being treated. Additionally, be aware of any medical conditions that may also impact mood (e.g., hypothyroidism). If not already discussed, capture when it was first discovered.

Important Coding Issues:

- **Education** – "Part College" = completed at least one year of college. Trade schools (hair dressing, auto mechanics, etc.) and certificate programs do not count as college.
- **Work History** – Do not be limited by the past 6 months. Get a snapshot of the stability of their work history. If a Respondent is a student, ask about grades and any changes in grades. (If presently a C student, is that typical?)
- **Inability to Work or Go To School** – If a Respondent indicates that they are on disability, confirm what the disability is (mental or physical), and when it was first diagnosed. Also ask how this disability impacts them today (symptoms/functioning).
- **Psychopathology** – Any treatment, even a single session, would be coded 3. Use the "Life Chart" page to capture the chronology of treatment. Since it is not always understood as treatment, ask specifically if s/he has ever received medication to help their mood, nerves or to help them sleep. Note: treatment for drug and or alcohol use is NOT coded here.
- **Most-upset/Feeling the best** – Do not take "Don't know" for an answer. Pause and encourage.

  Best – is a clue for possible mania.
- **Alcohol/Drug use** – After assessing past year use, also get an idea of life-time use. (e.g., "When were you using the most?" "Have you ever used?")
- **Current Social Functioning** – Important for assessing the GAF and will require follow-up questions. Does current functioning represent a change? If no or few friends are reported, why? Is this typical or a change in functioning? How does she feel about this?

II. Screening Module

The Overview is followed by a set of nine screening questions. These questions are taken from the body of the SCID and are typically the initial questions asked by the SCID for the disorders being screened. These screening questions may help reduce the potential effects of a "negative response bias" that may be especially problematic in the later sections of the SCID.

Because of the structure of the SCID, there is a tendency for the respondent to notice that a 'yes' answer to the initial probe question in a section results in follow-up questions, whereas a 'no' answer results in a skip to the next section. This leads to some respondents giving 'no' answers as a way of speeding the interview along. By asking these questions up front, and using answers to these questions in the determination of whether a section should be skipped, response bias may be minimized.

In the Screening Module, you are to ask the questions without any follow-up elaboration. You will have the opportunity to ask the respondent additional follow-up questions later, in the section of the SCID in which the disorder is being considered. Therefore, when the respondent gives a positive response to one of these screening questions, you should explain, "We'll talk more about that later." A 'no' answer should be coded as a 1, an equivocal response as a 2, and a 'yes' response should be coded a 3. All screening questions should be asked, answered, and coded before proceeding further in the SCID.

Important Coding Issue:

- If you should decide to go into a section based on information obtained later in the interview despite an answer of "No" (coded 1) here, be sure to change the Screening code to either a 2 or 3.

III. Module A. Mood Episodes

This module begins with ratings for Past Year and Lifetime Major Depressive Episodes, followed by Past Year and Lifetime Manic Episodes, and Dysthymic Disorder. Lifetime Major Depressive Episode is only assessed if the criteria for Past Year Major Depressive Episode are not met. Likewise, Lifetime Major Manic Episode is only assessed if the criteria for Past Year Manic Episode are not met. If the criteria for either Past Year or Lifetime Mania are met, then the Dysthymia section is NOT administered. Note that in this version of the SCID, Module A is the only module in which lifetime symptoms assessed. Also, remember that behavioral referents are important for all symptoms.

A. Major Depressive Episode

Criterion A:

Establishing the Minimum Two Week Duration: When you begin to ask about a possible Major Depressive Episode, first determine whether there has been a two-week period of depressed mood and/or loss of interest (either in the past year or before the past year). If there is some doubt about whether the duration of the depressed mood is truly two full weeks, inquire about the specific symptoms anyway, because it often turns out that a respondent who minimizes a problem when first asked may, on further reflection, recall that he/she was, in fact, symptomatic for a full two weeks. For example, let's say a respondent reports being depressed for several months during his junior year of college. Try to pinpoint a two-week interval as follows: "I know it's hard to be precise, but I need to focus on a two week period when it was the worst. Were you depressed during the fall semester of your junior year, or in the spring? Respondent answers "spring." Was it before or after spring break?" "How close was it to finals?" etc.

Establishing Co-occurrence of Symptoms during the Same Two-Week Period: In those situations in which the respondent reports more than one major depressive episode in the period in question, you should establish which of the episodes was "the worst" and subsequent questions should focus on the worst two-week period during that "worst" episode. It is recommended that you use holidays, seasons, or other life events (e.g., birthdays, graduation) as "landmarks" in trying to narrow down the two-week period in which the depression was the worst. The process of carefully reviewing the respondent's past also serves to transform the time period from an abstraction to a more vivid memory so that reporting of specific symptoms is more likely to be valid.

Bereavement/GMC/Substance Ruleouts: Do not assume that a period of depression may be complicated by either a death or drug use will "fall out" at either Criteria D (due to substance use/general medical condition) or Criteria E (bereavement). There are two reasons for this: 1. In the case of bereavement, if the symptoms lasted longer than 2 months or are characterized by marked functional impairment, morbid preoccupation with worthlessness, suicidal ideation, psychotic symptoms or psychomotor retardation, then the episode would be classified as a Major Depressive Episode (and not "Simple Bereavement"); 2. Even if the symptoms are better characterized as Simple Bereavement, or a Mood Disorder due to GMC/Substance Use, and a diagnosis of MDE is not given, the symptoms will nonetheless affect the GAF (for Past Year

episodes) and thus need to be reported on and documented. When a respondent reports multiple episodes of depression, some of which are complicated by Bereavement/GMC/Substance Use, consider first assessing the episode NOT complicated by these factors even if it was not the "worst" episode. If the respondent does not meet criteria for the "clean" episode, then go back and ask about the worst episode even if it is complicated.

Common Pitfalls: One of the most common errors in this section is the failure of the interviewer to ensure that each symptom was present for "most of the day, nearly every day." It is strongly recommended that you specifically ask, "Was that true for most of the day, nearly every day, during this period?" after each symptom, even to the point of tedium, since there is no other way to be sure this duration requirement is met. Do not assume that if the first several symptoms were present for most of the day, nearly every day, that the rest will also have this pattern. Note that item A(9) does not have to be present every day—recurrent suicidal ideation or a single instance of suicide attempt warrants a rating of 3. Further, all symptoms, including low mood and anhedonia, must be present during the SAME 2-week period.

A second common pitfall is to forget to ask about the second half of an item when the respondent answers "no" to the first part. For example item A(7) can be rated 3 if EITHER worthlessness OR excessive or inappropriate guilt has been present. If the respondent denies feelings of worthlessness, then do NOT code the item a 1 but follow-up that question with an inquiry about excessive or inappropriate guilt.

A third issue is how to count symptoms that occur in the context of a co-morbid general medical condition. General medical conditions may present with the same symptoms as characterize a depressive episode (e.g., weight loss, insomnia, fatigue). Should they be attributed to the depression or the medical condition? The rule in DSM-IV is to consider such symptoms as part of the depressive episode UNLESS they are clearly attributable to the medical condition. For example, insomnia related to frequent nocturnal coughing spells in a person with bronchitis should not count toward A(4). Similarly, symptoms should not count toward Major Depressive Episode if they are better accounted for by another disorder. For example, weight loss due to refusal to eat food because of the delusion that the food is poisoned should not be rated a 3.

A final issue is whether to consider as part of a depressive episode symptoms that have been present prior to the onset of the episode (e.g., chronic insomnia). Such symptoms count toward a diagnosis of Major Depressive Episode only if they have become appreciably worse

during the depressive episode. For example, if an individual who usually takes 30 minutes to fall asleep finds that it has been taking two hours to fall asleep since the episode began it would make sense to rate item A(4) as present for the episode.

Criterion A(1): Depressed mood may be acknowledged directly ("I've been feeling depressed") or by one of its many synonyms (sad, blue, tearful, "down in the dumps," "I can't stop crying"). Depressed mood in a Major Depressive Episode can be distinguished from "ordinary" (i.e., nonpathological) depression only by virtue of its persistence and severity. To count toward this criterion, the respondent's depressed mood must have been present for most of the day, nearly every day, for at least two weeks. Note that depressed mood in an adolescent may present as irritable mood.

Criterion A(2): Although the cardinal symptom of a Major Depressive Episode is depressed mood, it may be diagnosed in the absence of a subjective feeling of depression. Some respondents, particularly those with severe presentations, have lost the capacity to feel sadness. Others may have a cognitive style or come from a cultural setting in which feelings of sadness are downplayed. For such respondents, loss of interest or pleasure counts as a "depressive equivalent" and can be substituted for depressed mood when defining the two-week interval that applies to criteria A(3)-A(9). Evidence of this symptom may be that the respondent reports a general diminishing of pleasure (e.g., nothing makes me happy anymore) or specifies examples such as no longer reading books, watching TV, going to the movies, socializing with friends or family, or having sex. When rating this item, note that complete loss of interest or the ability to experience pleasure is not necessarily required for a rating of 3—evidence that there is a significant reduction in the ability to experience pleasure will suffice. Note: a loss of interest that is not due to mood is not coded 3. For example, a loss of interest in playing soccer because the respondent is now playing tennis or because he moved and that sport is not available, or loss of interest in gardening because it is now winter (too cold for gardening), would not qualify as anhedonia.

Criterion A(3): This item is rated 3 if there has been a significant change in appetite, either up or down, OR a significant change in weight during the two-week target period. Note that the first part of this item focuses on appetite and not on the amount of food consumed; thus a rating of 3 should be made only if the respondent acknowledges a significant change in his/her appetite for food. Since this is a compound item requiring a change in either appetite or weight, one need inquire about weight change only if there is no significant change in appetite. (Be

forewarned, however, that significant changes in weight without corresponding appetite changes suggest the possibility that a general medical condition may be responsible for the change in weight.) Note: weight gain or loss is over a one month period (not 2-weeks).

Criterion A(4): Insomnia may be manifested in many different ways, any one of which can count for this item. These include difficulty falling asleep, waking up a number of times in the middle of the night, and awakening much earlier than is normal for that person, with an inability to fall back asleep. Hypersomnia is sleeping much more than is normal for the person. Note that it is difficult and potentially not very meaningful to establish an absolute definition of the number of hours of sleep that constitute insomnia or Hypersomnia because of wide variability in individuals' need for sleep. However, as a rule of thumb, sleeping two hours more or less than is typical on a daily basis would constitute Hypersomnia or insomnia. Note that Hypersomnia should not be coded for someone who stays in bed for most of the day but is not sleeping. Also, if a respondent is taking a sleep medication, and thus sleeping ok; for coding, do not assume that the respondent would have a sleep problem if he/she were not on the medication. So, do not code it a 3 unless the respondent has actually reported sleep troubles during the episode.

Criterion A(5): Psychomotor agitation and retardation refer to changes in motor activity and rate of thinking (i.e., noticeably slowed or rapid speech and/or thinking and/or slowed or restless body movement). While many depressed respondents describe a subjective feeling of being restless or slowed down, this item should not be counted unless the symptoms are visibly apparent to an outside observer (e.g., the subject is either pacing and unable to sit still or seems to move in slow motion). There must be a convincing behavioral description of past agitation or retardation that was observed by others. Distinguish the feelings of being slowed down by psychomotor retardation (e.g., "I feel like I'm walking through a vat of molasses) from feelings of having no energy, which are coded in the next item.

Criterion A(6): Respondents with this symptom may report feeling tired all the time, "running on low power," "feeling weak," or totally drained after minimal physical activity. When a respondent complains about not feeling like doing anything, differentiate between lack of energy and loss of interest or motivation, which may also be present.

Criterion A(7): Be careful in rating this item since subjects who are depressed but do not have the full syndrome of Major Depressive symptoms often acknowledge feeling bad about themselves or feeling guilty. The actual item requires a more severe disturbance in self-

perception – either feelings of worthlessness OR excessive or inappropriate guilt. If, after asking a respondent how he/she feels about him/herself, you get an "I feel bad" or "I don't like myself," it is often helpful to present the respondent with the actual item (e.g., "How bad does that get? So bad that you feel worthless?") The worthlessness should be fairly global, and not just related to a precipitating stressor (e.g., if a breakup with a significant other precipitated the depression, worthlessness that is only related to the respondent's perception of herself as a girlfriend would not be coded as a 3). Similarly, while respondents often report feeling guilty about the negative impact their problems have on others ("I feel so guilty to be such a burden."), such feelings are often not excessive or inappropriate. A true positive response requires evidence of exaggerated and inappropriate guilt (e.g., "I feel like I've ruined my family forever").

Criterion A(8): Cognitive impairment in depression is sometimes severe enough to resemble dementia. With less severe, but still significant, impairment, an individual may be unable to concentrate on any activity (e.g., watching TV, reading a newspaper) due to an inability to filter out brooding thoughts. Interviewers should note that the impairment caused by this symptom may vary depending on the individual's baseline. For example, a theoretical mathematician may still be able to watch TV but no longer be able to concentrate on mathematical proofs—in such an instance, a rating of 3 would be warranted. Note that the second half of this item taps a different type of impairment (i.e., indecisiveness). A respondent suffering from this symptom may report feeling paralyzed by even simple decisions, like which clothes to wear for the day or what to eat for lunch.

Criterion A(9): This is the only symptom that does not have to be present nearly every day for at least two weeks to warrant a rating of 3. Any recurrent suicidal thoughts or behavior or any single suicide attempt is sufficient, as well as frequent thoughts such as: "I'd be better off dead" or "My family would be better off if I were dead." If there are current suicidal thoughts, it is imperative that you explore the seriousness of these thoughts and takes appropriate action. If a respondent reports any such thoughts, it is imperative that you ask when the respondent last had these thoughts, and especially if any occurred in the past two weeks (see Distressed Respondent Protocol in ***Section 5.7.4***). Self-mutilating behavior (cutting, burning, etc.) without suicidal intent that is only an expression of anger or frustration or is done with the aim of controlling anxiety, is coded 1. Chronic passive thoughts of death that are no worse during this period than any other would be coded 2. If this symptom is coded 3, then the GAF must be at 50 or below (serious symptoms).

Criterion C: Clinical Significance: DSM-IV has added this "clinical significance" criterion to most disorders to emphasize the requirement that a symptom pattern must lead to impairment or distress before being considered a diagnosable mental disorder. Do not assume – Ask!

Criterion D: This criterion instructs you to consider and rule out a general medical condition or a substance as an etiological factor.

Criterion E: Ruling Out Simple Bereavement: At this point in the SCID, a depressive episode lasting at least two weeks has been identified. If, however, that depressive episode is considered to be a normal and expectable reaction to the death of a loved one, it is not diagnosed as a Major Depressive Episode, but would instead be Bereavement. Therefore, the judgment required in this item is whether the depressive reaction has crossed the line (in terms of persistence and severity) from Bereavement to Major Depressive Disorder. DSM-IV provides two guidelines for diagnosing Major Depressive Disorder rather than Bereavement: 1) if the depressive symptoms (i.e., at least 5, most of the day, nearly every day) have persisted for at least two months following the loss, or 2) if certain depressive symptoms that are particularly uncharacteristic of Bereavement are present (e.g., suicidal ideation, psychomotor retardation, etc.). Note that this is one of those double negative items, ("Yes, we have NO Bereavement"), and a 1 is coded if it IS Bereavement.

Coding Note: A12/A13 ("PAST YEAR MAJOR DEPRESSIVE EPISODE STARTED:" Record the month and year that the depressive episode started, NOT when the 2-week period that was being assessed began.

B. Manic Episode

The issues for rating Past Year Manic Episode and Lifetime Manic Episode are the same as with Past Year and Lifetime Major Depressive Episode. Remember that if criteria are met for Past Year Manic Episode, there is no need to ask about Lifetime Manic Episode. Also recall that if in those situations in which the respondent reports more than one manic episode in the period in question, establish which of the episodes was "the worst" and subsequent questions should focus on the worst period during that "worst" episode. Use holidays, seasons, or other life events (e.g., birthdays, graduation) as "landmarks" in trying to narrow down the two-week period in which the depression was the worst. The process of carefully reviewing the respondent's past also serves to transform the time period from an abstraction to a more vivid memory so that

reporting of specific symptoms is more likely to be valid. For all symptoms, get behavioral referents. The symptoms during this period must present differently from "normal."

Criterion A: Elevated or Irritable Mood: This criterion requires a persistently elevated, expansive or irritable mood. Respondents often describe periods of irritability that are clearly not associated with a Manic Episode. Most commonly, such periods are either Major Depressive Episodes with irritability as an associated feature, or chronic irritability that is a symptom of a personality disturbance. However, if there is any question whether the irritability might be part of a Manic Episode, continue to ask all the manic questions in order to make a judgment as to whether the irritability is a symptom of a Manic Episode or is better accounted for by another condition.

Criterion A: One Week Duration: The criteria set for Manic Episode has a minimum duration of one week.

Criterion B: For all of the B items, it is imperative that examples be obtained. The examples should be quite convincing. For each symptom, it is useful to assess whether other people noticed the symptom and commented on it. While the symptom does not *need* to have been commented upon to warrant a 3, such comments do bolster confidence in the rating.

Criterion B(1): It is important to remember that in order to count a B symptom toward a diagnosis of a Manic Episode, the symptom must be present during the period of elevated or irritable mood and must be persistent and clinically significant. Therefore, merely being more self-confident than usual would not warrant a rating of 3. There must be grandiosity or inflated self-esteem that is clearly not justified by a realistic evaluation.

Criterion B(2): The respondent should report getting by on 2 (or more) hours less sleep than usual in order to justify a rating of 3 on this item. The prototypic individual feels that he or she does not need sleep at all. While some individuals may report waking up and not feeling tired, more commonly the individual describes feeling very driven or "wired" and cannot calm down enough to sleep. It may be useful to probe what they are doing when they are not sleeping.

Criterion B(3): The increase in talkativeness is manifested in both the rate and amount of speech. The speech often has a driven quality, as if there is so much to say and not nearly enough time to say it. If present during the interview, it may be very difficult for you to interrupt the respondent's monologue. It may be useful to ask the respondent whether other people *commented* that they had trouble interrupting the respondent or getting a word in edgewise.

Criterion B(4): Flight of ideas involves thoughts that are loosely connected, with the respondent jumping from one topic to another very quickly, with only the slightest thread of thematic connection between topics. In some cases, the connection may be based on sound rather than meaning (clang association). It is useful to use the phrase "racing thoughts" and ask the respondent whether he/she would describe the thoughts as "racing."

Criterion B(5): Distractibility refers to an inability to filter out extraneous stimuli while attempting to focus on a particular task. For example, the respondent may have trouble focusing on the questions because of being distracted by a police siren on the street, and may need to jump up from the interview to investigate what is going on outside. The extraneous stimuli should be *external* and *irrelevant*.

Criterion B(6): As a consequence of elevated mood, increased energy or increased self esteem, the person may become involved in more activities than usual. Typical "manic" activities involve calling friends at all hours of the night, writing lots of letters, beginning new creative projects. Alternatively, the increase in activity may be more diffuse and be manifested as psychomotor agitation (e.g., being unable to sit still). Similar to the psychomotor agitation in the MDE section, other people must have noticed and commented on the agitation in order to warrant a rating of 3.

Criterion B(7): In the pursuit of pleasure, excitement or thrills, the person may engage in activities that are uncharacteristic of them, without regard to possible negative consequences. Typical examples include lavishly spending large sums of money on luxury items, gifts for others, or expensive vacations, driving too fast, or engaging in reckless or unsafe sexual behavior. Note that the SCID question (i.e., "did you do anything that could have caused trouble for you or your family?") is relatively non-specific in that it will pick up any behavior that reflects poor insight or judgment. This item should be coded 3 ONLY if the behavior is extremely pleasurable and the respondent is acting without regard for the potential painful consequences. For example, problematic behavior occurring in response to a command hallucination or delusion (e.g., accosting strangers with the news that doomsday is approaching) would not warrant a rating of 3.

Criterion B (3/4 out of 7): Note that the number of items required to meet criterion B depends on whether criterion A was coded 3 in absence of euphoric mood (i.e., irritable mood

only). If euphoric mood has been present, then only three B criteria need to have been present. Irritable mania requires a minimum of four items to help differentiate it from irritable Major Depressive episodes. Make a note in the SCID if this item is coded 1 with three symptoms present, but mood is irritable.

Criterion D: As indicated in this criterion, the symptoms in a Manic Episode must be sufficiently severe so as to cause marked impairment, require hospitalization, or include psychotic features.

Criterion E: This criterion instructs you to consider and rule out a general medical condition or a substance as an etiological factor.

C. Dysthymic Disorder

Criterion A: The purpose of this item is to determine whether or not the respondent has had a depressed mood for more than 50% of the past two years (or for the two years before the date an MDE occurred in the past year). If a respondent reports being depressed for a significant portion of her life, a depressed state may seem "normal" and therefore go under-reported. Be alert for this type of situation and use additional probes as necessary.

Criterion B: Items 2, 3, 4 and 5 are identical to the corresponding items for Major Depressive Episode, except that they need not occur nearly every day for at least two weeks. All of the symptoms may be intermittent, but must be present more than half of the days in the two-year period under consideration. Item 4 is set at a lower threshold, in that it requires only low self-esteem, and not a feeling of worthlessness or inappropriate guilt. Be sure to obtain behavioral referents for each physical symptom even if MDE has been assessed. Also, be watchful for discrepancies between report of MDE symptoms and those reported here. Note that all symptoms must be present more than half of the days in the two-year period under consideration when depressed, and not present or significantly different at other times.

Criterion C: In order to ensure the chronic nature of Dysthymic Disorder and to differentiate Dysthymic Disorder from recurrent minor depression, make sure the respondent never had any significant (i.e., more than two months) depression-free periods. If the respondent gives a pretty convincing description of constant low-lying depression, and then says there have been 2 month periods w/o symptoms, be sure to ask questions around this (e.g., what were those symptom free months like? How was your mood? Sleep? etc.).

Criterion D: As discussed above, whether a chronic depressed mood represents Dysthymic Disorder or Major Depressive Disorder in Partial Remission depends on the severity of the symptoms and the initial onset of the depression. If the disorder begins with a two-year period of depression that is less severe than a Major Depressive Episode, the diagnosis of Dysthymic Disorder is appropriate. If no such two-year period can be documented, the depression is better accounted for by a diagnosis of Major Depressive Disorder in Partial Remission with a prolonged depressive prodrome.

Criterion F: This item rules out depressive periods that occurred exclusively during the course of a chronic Psychotic Disorder.

IV. Module B/C. Psychotic Screen

This module is used to determine whether a psychotic symptom has been present during the past year; but it does not include enough detail to make a complete differential diagnosis. Do not assume that an experience described is "normal." Probe and get details about how frequently it occurs, what meaning the respondent gives to the experience, and how common the experience is in the context of the respondent's culture. Be aware of cultural norms that are different from your own.

A. Rating Delusions

A delusion is a false personal belief based on incorrect inference about external reality that is fully sustained despite what almost everyone believes and despite what constitutes incontrovertible and obvious proof or evidence to the contrary. The belief is not one ordinarily accepted by other members of the person's culture or subculture (e.g., the belief in some cultures that one can communicate with a dead person). When you are unfamiliar with the beliefs characteristic of the individual's cultural or religious background, consultation with someone who is familiar with the respondent's culture may be required to avoid over-diagnosis of delusions.

A delusion involves impairment in the ability to make logical inferences—the way conclusions are drawn from observations of the person's environment or self (e.g., phone hang-

ups indicate that the person is being spied on). In rating each type of delusion, you must differentiate a delusion (which would warrant a rating of 3) from a strongly held "overvalued" idea (which would warrant a rating of 2). In deciding whether a belief is false and fixed enough to be considered a delusion, first determine that a serious error in inference and reality testing has occurred and then determine the strength of the conviction. It may be helpful to ask the respondent to talk at length about his/her conviction because it is only in the specific details that the errors of inference become apparent. In evaluating the strength of the delusional conviction, the interviewer should present alternative explanations (e.g., is it possible that the phone hang-ups are due to people dialing a wrong number?). A delusional respondent may acknowledge the possibility of these explanations, but still hold firm to his/her own belief. Some respondents with a longstanding history of psychotic disorder have developed insight into the "psychotic" nature of their delusions. Such a symptom would still be considered "psychotic" as long as, at some earlier point, the symptom was experienced as real. For example, a respondent may report that his chronic conviction that people at work are plotting against him is a result of his longstanding Schizophrenia. This would be coded as a delusion if the respondent reports that initially he was convinced the plot was real.

The next set of ratings documents the type of delusion, based on content. Note that for a particular delusion, more than one rating may apply. For example, a respondent who believes the FBI is after him because he can control other people's minds would have both persecutory and grandiose types of delusions coded 3.

Delusion of Reference: This question has a relatively high false positive rate. Ask for specific examples that establish the psychotic nature of the belief. Most people have at some time felt that other people were talking about them, particularly if they have some obvious physical abnormality or act in a way that makes them stand out. It is therefore important to differentiate realistic perceptions, social anxiety or transient suspiciousness from a fixed false belief. A homeless man who dresses in rags and has no place to take a shower may realistically believe that people are moving away from him on the subway, but if he believes that today's headlines are a cryptic reference to his personal life, the interviewer should rate this item a 3.

Persecutory Delusion: Take care to differentiate an exaggerated, possibly valid, perception (e.g., by a boss, a teacher, an ex-spouse, a drug dealer) from real persecutory

delusion. There may be cases in which it is impossible to know whether the persecution is real or delusional.

Grandiose Delusion: It is sometimes hard to tell where an inflated perception of one's talents ends and where a grandiose delusion begins. A taxi driver who believes he will write a best-selling novel may be mistaken, but not necessarily delusional. If, however, he tells the interviewer that Steven Spielberg has been calling, begging for movie rights to his novel, he has probably stepped over the line into delusion. Questioning him about his evidence for the belief is a good way to clarify the issue.

Somatic Delusion: In assessing this symptom, it is necessary to take into account the respondent's understanding of anatomy and physiology. An uneducated person may have a primitive explanation of symptoms, for example, believing that stomach pains are caused by a grasshopper hopping around inside of him. His willingness to entertain an alternative explanation indicates that the belief is not a delusion. Another example of a false positive would be a respondent with physical symptoms who doubts an internist's reassurance that she has no medical illness. If the respondent is willing to entertain the possibility that her beliefs are exaggerated, then the diagnosis would be Hypochondriasis. A respondent who dismisses such reassurances out of hand is more likely to have a somatic delusion.

Other Delusions: Other types of delusions that do not easily fit into one of the specified types are coded here.

B. Rating Hallucinations

Auditory Hallucinations: Auditory hallucinations should be differentiated from delusions of reference, in which the respondent hears actual voices (on the street, etc.) and interprets then self-referentially. Evidence that they are, in fact, hallucinations might be that they occur when the respondent is alone. This item should be coded 3 only if hallucinations are judged to be clinically significant, (i.e., recurrent or persistent). Hearing one's name being called and finding no one there is an example of a hallucination that is not clinically significant. A hallucination (the experience of a sensory perception without stimulation of the relevant sensory organ) should also be distinguished from an illusion, which is a misperception of an actual stimulus (e.g., misinterpreting a shadow as the figure of a man).

V. Module D. Mood Disorders

This module is for recording Mood Disorder Diagnoses (other than Dysthymic Disorder, Mood Disorder Due to a General Medical Condition, and Substance-Induced Mood Disorder). The diagnoses covered in this module include Bipolar I Disorder and Major Depressive Disorder. The task in this module is to evaluate whether the specific criteria for Mood Disorders are met based on information gathered in Module A. If any symptom in either the MDE or Mania sections is coded 3, then D1 would also be coded 3.

VI. Module E. Anxiety Disorders

A. Posttraumatic Stress Disorder (PTSD)

The evaluation of PTSD begins with a screen that first reviews the individual's lifetime history of exposure to severely traumatic experiences and then determines whether any of these traumatic experiences have been re-experienced in the form of dreams, flashbacks, intrusive thoughts, or strong reactions when in situations that are reminiscent of the trauma. If so, the interviewer should proceed with the evaluation of PTSD, focusing on the event identified in the screen. If more than one traumatic experience is reported, ask the respondent to choose the one that seems to have affected him/her the most during the past year. If, during the evaluation of PTSD for this particular stressor, it becomes clear that the criteria for PTSD are not being met, determine whether one of the other stressors might in fact have had a greater impact on the individual during the past year, and then reevaluate the criteria in relation to this different stressor.

Criterion A(1): In evaluating whether a stressor qualifies as a potential source of PTSD, both the type of stressor ("actual or threatened death or serious injury or threat to the physical integrity of self or others") and the context of exposure ("experienced, witnessed, or confronted with") should be considered. In DSM-IV, stressors are limited to events that pose a threat to life, limb, or physical integrity. Stressors that, while distressing, are not life-threatening (e.g., being humiliated by a boss at the office) do not warrant a rating of 3 for this item. Although the prototypic stressor for PTSD is a wartime combat experience, the concept has been expanded to include other life threatening experiences like being the victim of a serious crime, accident, or disaster. The phrase "threat to physical integrity" includes all experiences of sexual assault or sexual molestation, not just those in which the victim perceives a threat of violence. The context

of the exposure includes having one's life threatened; having the direct personal experience seeing someone else being threatened, injured or killed; or hearing the news of a loved one being hurt. It is not meant to include more indirect and impersonal experiences such as hearing a news report of a catastrophe occurring to strangers. Similarly, the expected death of a loved one of natural causes at an advanced age does not qualify as a PTSD stressor.

It is very important to get a description of the event and some details of what happened. Be respectful/show empathy, but don't assume that the respondent cannot talk about it. Do not begin by giving the respondent permission to not answer (e.g., "you do not have to answer these questions if you don't want to"). While it is true that respondents do not need to answer anything they are uncomfortable answering, you should administer the module with the assumption that the respondent will answer. Getting some details will be essential in assessing the Criteria B, C, and D.

Criterion A(2): This criterion requires that the person be profoundly affected by the stressor and react to it with extreme feelings of "fear, helplessness, or horror." It is possible that the respondent says he/she had no emotion initially; he/she may have been in shock. If this is the case, assess if he or she felt differently as a later time.

Criterion B: The re-experiencing of the traumatic event can occur spontaneously (intrusive memories, flashbacks, or dreams) or can be triggered by a wide variety of stimuli that remind the person of the traumatic event. For example, smoke from a campfire may produce profound terror in someone who has been trapped in a house fire. Flashbacks to wartime may be triggered by loud noises, seeing war movies, or tropical rainstorms. In some cases, the trigger may be a symbolic representation of the actual stimulus (e.g., a policeman for a concentration camp survivor). Occasionally, it may be difficult to differentiate a flashback from a psychotic experience. In contrast to psychotic symptoms, the sense that one is reliving a traumatic experience in PTSD is transient, self-limited, and understandable in the context of the exposure to the prior stressor. Note that the re-experiencing must be persistent—a few nightmares or intrusive memories don't satisfy this criterion. Therefore assessing the frequency for each symptom is essential. Clinical judgment must be used in assessing significance; however, a symptom that is present less than monthly would not likely warrant a rating of 3. Further, the cues or triggers for both emotional distress and physiological reactivity must be linked to the traumatic event. It is important to elicit examples of the symptoms and to ask additional probes

as necessary. Remember, for B(1), the thoughts must be intrusive/distressing, not just memories of or thoughts about the trauma. For B(2), be sure to ask about the content of the dream. Again, it must be distressing (not just "another dream"). The dream must also be related to the traumatic event. It is important to note that B(3) is a rare symptom – flashbacks are not merely "strong" memories of the event. There must be some sense of reliving the experience. Finally, for B(4) and B(5), information must be gathered regarding what the cue or trigger was and how the respondent reacted. A preference to not watch the TV news because of reporting of similar events would not be coded 3.

Criteria C and D: These criteria include symptoms that are much less specific than those in A and B, and are seen in many other disorders. Many people, for instance, try to avoid talking or thinking about bad things that have happened to them, whether or not the bad things were traumatic. Diminished interest, detachment, a restricted range of affect, insomnia, difficulty concentrating, etc., may be symptoms of a Depressive Disorder or of a Personality Disorder. It is important to clarify that the symptoms in C and D developed after the trauma. (In the case of a childhood trauma, it is impossible to know what the person would be like had he or she not had the experience. Following the SCID principle that one does not make a diagnosis without the evidence, we would be hesitant to diagnose PTSD in an adult who has had symptoms as long as he or she can remember.) Again, the symptoms must be present in the past year, have developed after experiencing the trauma and be persistent since that time. In terms of gathering additional information, remember that for C(2), the person/thing being avoided must be directly related to the traumatic event. For C(4), recall that, as in anhedonia, the diminished interest must not merely represent a change of interest or decrease of interest due to other factors.

Criterion F: This criterion instructs you to assess if the symptoms are clinically significant.

B. Panic Disorder

Prior to assessing the presence of Panic Disorder, have the respondent describe in detail what happens when he/she has a panic attack.

Criterion A(1): The term "panic attack" is often incorrectly used by respondents to describe any escalating anxiety, but the hallmark of a true panic attack is the sudden and intense onset of symptoms. The physical symptoms and terror are often overwhelming, and initial panic

attacks may lead a person to seek emergency care because of concerns that he/she may be having a heart attack. Since the term "panic attack" is liberally used, make sure you have the respondent describe what they mean by this expression prior to assessing any of the criteria.

The presence of a panic attack is not necessarily indicative of Panic Disorder since panic attacks can occur in the context of a number of Anxiety Disorders. For example, if a person with a snake phobia goes on a hike and has a panic attack after accidentally stepping on a snake, this would not warrant an additional diagnosis of Panic Disorder. By definition, at least two of the panic attacks in Panic Disorder must have been "unexpected." Only one of the panic attacks must have occurred in the past year.

Assessing whether a panic attack was "unexpected" may be difficult because respondents with Panic Disorder commonly (and mistakenly) believe that there is a cause-and-effect relationship between the situations in which the attacks have developed and the attacks themselves. For example, a respondent who experienced several unexpected attacks while shopping may assume that it was the experience of being in a crowded store that led to the attacks and therefore not consider the attack to be "unexpected." In such a situation, use the follow-up question, "Have you ever had a panic attack when you didn't expect to at all?" or "When you were in the store, right before you had your first attack, were you already feeling anxious or were you feeling okay?" For some individuals, attacks occur following a frightening thought, such as worrying that something will happen to them or a loved one. Such attacks should still be regarded as "unexpected" because the concept of "unexpected" refers to the absence of a clear association between an environmental stimulus and the occurrence of the panic attack. Common sense will lead you not to include as "unexpected" panic attacks that occur in response to unexpected but realistic dangers, such as being mugged. Similarly, panic attacks that occur in response to delusions about being harmed should not be regarded as "unexpected."

Criterion A(2): The diagnosis of Panic Disorder requires that at least one of the panic attacks be clinically significant. The three subparts of this criterion present three different ways in which clinical significance may be manifested: either persistent worry about the implications of the attack (e.g., that the attack means the respondent is going crazy or suffering from a medical illness), worry about having additional attacks, or a change in lifestyle (most commonly, avoidance of situations or activities or modifications to accommodate the attack, like always

being near an exit, sitting on the aisle of a theatre, etc.). Respondents who have had panic attacks but are not particularly bothered by them would warrant a rating of 1 for this item.

Criteria for Panic Attack: In DSM-IV, panic attack is defined using a free-standing criteria set that is not part of the criteria set for Panic Disorder per se. The SCID embeds the evaluation of whether the discrete episode of symptoms constitutes a true panic attack or, rather, just a limited symptom attack (i.e., less than 4 symptoms) in the evaluation of Panic Disorder. Before presenting the 13 symptoms of a panic attack to the respondent, the interviewer is instructed to first inquire about the most recent panic attack in an open-ended way (i.e., "when was the last bad one? What was the first thing you noticed?" etc.) This allows the respondent an opportunity to describe the attack in his or her own words, which is helpful in determining whether the experience has the qualities of a true panic attack. This method is also helpful in encouraging the respondent to focus on a single specific attack when endorsing the characteristic symptoms. Note that if fewer than four symptoms are endorsed, the interviewer should ask if the respondent has had other attacks in the past year during which there were more symptoms. If so, the interviewer should go back and reapply the list of 13 symptoms to this most severe attack and then confirm that such attacks have occurred repeatedly.

This requirement that the symptoms reach a peak within ten minutes is to differentiate a panic attack from slowly escalating anxiety. In fact, panic attacks often peak within seconds, and almost always within a few minutes. Although most panic attacks subside within an hour, some individuals may continue to have symptoms and a high level of anxiety for hours after the peak.

Criterion C: This criterion instructs you to consider and rule out a general medical condition or a substance as an etiologic factor for the panic attacks. Remember to carefully assess caffeine intake, and remember that caffeine is present in a variety of foods and over-the-counter medications like Anacin. Although substance use may be associated with the initial onset of panic attacks, a substance use etiology should be considered when subsequent panic attacks occur ONLY in the context of substance use.

Do not pre-judge the impact of substances or medical conditions. Caffeine and other stimulants, hypoglycemia and hyperthyroidism all can bring on similar symptoms (a more complete list is in the SCID). For some people *any* amount of caffeine can bring on panic-like symptoms.

Merely asking if the respondent sees a connection between the symptoms and these other factors is not sufficient to code this item. A temporal and consistent relationship must be established.

Criterion D: This criterion asks you to consider whether the panic attacks are better accounted for by another mental disorder. This judgment depends on determining whether the panic attacks are cued by an anxiety-provoking stimulus arising in the context of another disorder. For example, consider an individual with longstanding Social Phobia who has a panic attack while speaking in front of a large group of people. Since the panic attack was triggered by exposure to an anxiety-provoking situation (e.g., speaking in public) it is considered to be better accounted for by the Social Phobia. Similarly, if someone with Posttraumatic Stress Disorder develops a panic attack when exposed to a stimulus that reminds the person of the traumatic stressor, then the attack would not be considered a symptom of Panic Disorder. Note that this criterion does NOT set up a strict hierarchy between Panic Disorder and other mental disorders (i.e., DSM-IV does not instruct you to never diagnose Panic Disorder if the attacks occur during another disorder). Instead, clinical judgment is required to determine whether the other disorder "better accounts for" the panic attacks. You may need to set aside the evaluation of this criterion pending completion of the remainder of the SCID, if it seems likely that another disorder which accounts for the panic attacks may be present.

Criterion B: Agoraphobia: Some individuals with Panic Disorder manage to grit their teeth and suffer panic attacks without developing any avoidance, but most begin to associate certain situations with their panic attacks and consequently avoid those situations (or else endure them only with great anxiety). This avoidance may range from simply not driving a car because a person is afraid of having a panic attack while driving, all the way to never leaving home because one is afraid of having an attack in a place that's not "safe." The key here is fear of a situation where help may not be available *in the event of a panic attack or panic like symptoms.* The fear is of having the panic-like symptoms, not of the situation.

Criterion B(1): To rate this item 3, the respondent should report that a situation is avoided because of fears that a panic attack is more likely to develop in that situation or that escape from the situation may be difficult or embarrassing in case of having a panic attack. In some cases, the respondent may not be aware of the reason certain situations are avoided. If the avoidance develops soon after the onset of panic attacks, you may infer that the avoidance is related to the panic attacks.

Criterion B(2): Note that a rating of 3 can still be appropriate for a respondent who is able to force him/herself to go into the agoraphobic situations so long as there is either marked distress or the need for a companion to accompany the person. Really pull for the clinical significance of the avoidance and effect on functioning. A respondent who leaves the house everyday to go to work, but who feels uncomfortable in crowds would not be considered agoraphobic.

Criterion B(3): This criterion is similar to Criterion D in Panic Disorder in that it reminds you to consider whether the fear and avoidance may be better characterized as part of another mental disorder. Two of the most difficult boundaries are with Specific Phobia and Social Phobia. Typically, Agoraphobia involves avoidance of a cluster of situations, reflecting the general unpredictability of panic attacks. In contrast, Specific Phobia tends to be limited to one consistently-feared situation. Furthermore, the onset of Agoraphobia is related to the onset of panic attacks, whereas Specific Phobias tend to be either lifelong or related to a traumatic experience. Determining whether avoidance of social situations is related to Social Phobia or to fear of developing a panic attack in a social situation (which would warrant a diagnosis of Agoraphobia) generally depends on determining the temporal relationship between the onset of panic attacks and the social avoidance. If an individual develops social avoidance only AFTER the onset of panic attacks, then Agoraphobia is the most appropriate diagnosis. An individual with longstanding social avoidance who develops panic attacks when in social situations would better be considered to have Social Phobia. Note that this criterion does NOT preclude making a diagnosis of BOTH Panic Disorder with Agoraphobia and another disorder characterized by avoidance in the same individual (e.g., an individual with a long-standing dog phobia since childhood who develops unexpected panic attacks in situations without the presence of dogs).

C. Agoraphobia without History of Panic Disorder

This disorder is analogous to Panic Disorder with Agoraphobia, but the concern is about panic-like symptoms as opposed to full-blown panic attacks. In Panic Disorder with Agoraphobia, there is anxiety about being in places or situations from which escape may be difficult or embarrassing in the event of having a *panic attack*. In this condition, the anxiety is focused on being in places or situations from which escape may be difficult or embarrassing in the event of

having *panic-like symptoms* (either a specific uncontrollable symptom like a loss of bowel control or else subthreshold versions of panic attacks known as "limited symptom attacks").

Criterion A: When assessing this criterion, it is important to establish that the anxiety and/or avoidance are related to concerns about having panic-like symptoms and not for other reasons (e.g., a person having anxiety about going outside because he/she lives in a dangerous neighborhood).

Criterion B: This criterion, which requires that criteria have never been met for Panic Disorder, has been omitted from the SCID since this criterion has already been met by virtue of the branching logic that skips this disorder if criteria have been met for Panic Disorder.

Criterion C: The type of anxiety or avoidance seen in this condition would rarely be associated with a substance or general medical condition (except as discussed in criterion D). Nonetheless, DSM-IV requires you to rule out these etiologies.

Criterion D: Individuals with general medical conditions often restrict their activities and avoid situations because of concerns that the symptoms of the general medical condition may be disabling or embarrassing. This criterion clarifies that this diagnosis may be appropriate if the amount of anxiety or avoidance is unreasonable given the severity of the general medical condition. For example, avoiding driving for several weeks following a severe heart attack would not warrant a diagnosis of AWOPD, whereas being housebound for two years following a mild heart attack might warrant a diagnosis.

D. Social Phobia

Criterion A: There is a wide range of social triggers that may qualify for this criterion—what they all have in common is that the respondent fears acting in a way that will be humiliating or embarrassing. Some people are afraid of any kind of scrutiny—they choose to work by themselves, won't go to parties, or out on dates because they are extremely self-conscious, and believe that others will judge them to be foolish, stupid, or inept. (This is often a lifelong pattern and these individuals usually also have Avoidant Personality Disorder.) Other socially phobic people are comfortable in many relationships, but uncomfortable about various situations in which they are required to "perform." This more circumscribed form of Social Phobia includes

traditional performance situations like public speaking or playing a musical instrument, as well as other behaviors that are only a performance in the person's mind: fears of urinating in a public bathroom, writing in front of others, or eating in front of others. Individuals with performance anxiety have no problem performing the behavior when alone (e.g., giving a speech in front of a mirror). In all cases, a rating of 3 requires that the focus of the anxiety is concern about being humiliated or embarrassed by the scrutiny of others. Avoidance of a behavior because of concerns that the person's own high standards will not be met (as in Obsessive-Compulsive Personality Disorder) would not warrant a rating of 3.

Because concerns about public speaking are so ubiquitous, it is important not to rate this criterion as a 3 for public speaking unless it is clear that the person's concerns are excessive and do not diminish with practice.

Remember, these need to be reactions that have occurred in the past year – not just sometime in the past. Also, be sure to get examples to support the criterion "marked and persistent fear."

Criterion B: The question *"Have you always felt anxious…"* is intended to assess the consistency of the reaction on every exposure to the feared situation; not an intermittent presence since childhood. This criterion should be rated 1 if the anxiety and avoidance are erratically expressed (i.e., fear of speaking in one class, but no fear of speaking in a different class with the same number of people).

Criterion C: This criterion helps to set the boundary between Social Phobia and the social avoidance that is characteristic of many psychotic disorders. In a psychotic disorder, the social avoidance is usually associated with a delusional belief (e.g., persecutory delusion) that the respondent believes justifies his/her social anxiety or avoidance.

Criterion D: This two-part criterion underscores that avoidance of social situations is not a required part of this disorder. A diagnosis of Social Phobia may also apply to those who force themselves to go to parties, give talks, or go on job interviews, but feel anxious while doing it.

Criterion E (also in Specific Phobia): A diagnosis of Social or Specific Phobia is not made unless avoidance, anticipatory anxiety, or distress is clinically significant (i.e., interferes with functioning, with social activities, or with relationships, or there is marked distress ABOUT

having the phobia). Thus, for example, a public speaking phobia in a plumber who is almost never called upon to address groups is unlikely to meet the criterion, as is a snake phobia in someone who rarely leaves New York City. Too, claims that promotions would have occurred had the respondent been more at ease socially (or better at public speaking) would not generally meet the criterion; the critical difference being the low level of skill or ease versus a true phobia. Some individuals who seriously constrict their lives to avoid social (or other phobic) situations may report a lack of distress since their phobias are never activated. A rating of 3 may still be justified if you make a clinical judgment that the phobia has a significant negative impact on their functioning. Note: Most potential diagnoses of phobias sink or swim on this criterion.

Criterion G: Phobias are rarely caused by the direct physiological effects of a substance or general medical condition.

Criterion H: DSM-IV excludes a diagnosis of Social Phobia for individuals with medical or psychiatric conditions who may understandably avoid social situations because they are embarrassed about their symptoms. Such individuals, especially those who overreact to their condition or disability, may deserve a diagnosis of Anxiety Disorder NOS (which would be recorded on page X.4 in the Interviewer Debriefing Module, if appropriate).

E. Specific Phobia

The criteria B, C, D, and E are the same as in Social Phobia.

Criterion A: The hallmark characteristic of a phobia is that the fear is way out of proportion to the degree of danger posed by the object or situation. Be sure to obtain good evidence of "marked and persistent fear that is excessive or unreasonable…" by eliciting behavioral referents.

Criterion G: Specific Phobia is, in a sense, residual to other disorders with stimulus triggered anxiety. For example, although fear and avoidance of contamination may meet criteria for a "dirt phobia," if the fear and avoidance occurs as part of a contamination obsession and hand washing compulsion in Obsessive Compulsive Disorder, then an additional diagnosis of Specific Phobia is not made.

It should be noted that a diagnosis of Specific Phobia can be made along with one of these other disorders if the fear and avoidance are unrelated to the other disorder. For example, a person with Panic Disorder may avoid many different situations or activities because of fear of a panic attack, but may also have specific phobias that are unrelated to the Panic Disorder. It is up to you to get enough information to judge whether, in addition to agoraphobia, there are fears (e.g., of dogs, of spiders) that are unrelated.

F. Obsessive Compulsive Disorder

Obsessions, Criterion A: The most common diagnostic problem is distinguishing true obsessions from other repetitive distressing thoughts, like excessive worries about realistic concerns, depressive ruminations, and delusions. Obsessions have an intrusive, inappropriate, and "ego-alien" quality and are experienced by the respondent as something different and stranger than the worries or preoccupations that characterize Generalized Anxiety Disorder or a normal reaction to life's unpredictability. The recurrent, intrusive, and anxiety-provoking thought while driving, that one ran over a small child without realizing it is an obsession.

Spending an equal amount of time worrying about one's retirement is more likely to be an aspect of Generalized Anxiety Disorder. Unlike obsessions, depressive ruminations and delusions are generally not perceived as intrusive or inappropriate, but are understood by the respondent as a valid form of concern, even if he/she realizes that the concern is excessive and tries to stop thinking about it.

In those situations when the differential diagnosis is particularly challenging, it may be useful to remember the fact that obsessions and compulsions usually go together (in fact, 90% of the time according to the DSM-IV OCD field trial). Therefore, in trying to distinguish between an OCD obsession and other repetitive thoughts, the clinching point may be whether compulsions are also present.

Important Interviewing/ Coding Issues:

- All four aspects of Obsessions (A1- A4) must be coded 3 in order to meet the diagnosis.
- A1 (E79) - these thoughts are more than recurrent - they must also be experienced as intrusive and inappropriate.

Compulsions, Criterion A: Compulsions are distinguished from other forms of repetitive behaviors by the underlying motivation for the behavior—to reduce or prevent the anxiety associated with an obsession. For example, hand washing alleviates anxiety triggered by the obsession that one is contaminated; repeating a prayer exactly 36 times is meant to counteract the distress caused by having an obscene thought. Determining that the behavior is intended to reduce the anxiety accompanying an obsession is very helpful in differentiating a compulsion from other repetitive behaviors like tics and stereotypes. The most common compulsions are behaviors like hand washing, repetitive touching or picking up and replacing an object over and over again, or mental acts such as counting or repeating a word or phrase over and over.

Important Interviewing/ Coding Issues:

- All four aspects of Obsessions (E79-E82) must be coded 3 in order to meet the diagnosis.
- E79 - these thoughts are more than recurrent - they must also be experienced as intrusive and inappropriate.
- E80 - worrying about safety when living in a rough neighborhood is not what we are looking for here
- E84/85 In assessing these items, be sure to ask the respondent what would happen if he/she were prevented from doing the behavior. Washing one's hands to prevent the spread of germs (especially in a health-care setting) would not typically qualify (useful probes: does the respondent do so more than others do? Is this standard of washing typical in this setting?)

Criterion B: This symptom (on page E.30) is to be assessed if the respondent meets criteria for either Obsessions or Compulsions. If the respondent meets criteria for both Obsessions and Compulsions, then these questions are to apply to both sets of symptoms. The prototypal individual with OCD is aware that his/her obsessions and compulsions are unreasonable (e.g., that it is ridiculous to be concerned about contamination from germs that might arise from touching newspapers). Over time, some individuals with OCD lose insight as to the excessive nature of their obsessive concerns or compulsive behaviors and may in a later stage of the illness describe the obsessions or compulsions as being reasonable.

For such individuals, it is essential to establish whether at some time in the past (usually early in the course of the illness) the obsessions and compulsions were regarded as unreasonable (e.g., when the hand washing first started, did you feel that you were washing your hands much more

than you should or than really made sense?). For those individuals who no longer recognize that the obsessions or compulsions are not reasonable, the specifier "With Poor Insight" may be noted.

Criterion C: This criterion requires that the obsessions or compulsions be clinically significant. Note the standard DSM-IV clinical significance criterion also includes a phrase indicating that the obsessions or compulsions may be "time consuming (take more than an hour a day)." This clause allows you to conclude that impairment is present even in the face of respondent's lack of concern about the behavior or the realization that it is useful.

Criterion D: An additional diagnosis of OCD should not be given along with another mental disorder if the repetitive thoughts or behaviors can be considered to be features of the other mental disorder. Most of the examples of symptoms of other disorders that are given in the SCID do not really meet the test of "intrusive and inappropriate." For example, when a patient with Anorexia Nervosa is preoccupied with measuring the exact number of calories in the food she eats, she may agree only that it is excessive, not foolish. However, if the obsessions or compulsions are clearly symptoms of another disorder, the interviewer may skip out of the diagnosis of OCD without spending a lot of time deciding whether the symptoms are "intrusive and inappropriate" or just "excessive." (Of course, Anorexia does not protect someone against OCD; the individual with Anorexia may also have hand washing rituals that are unrelated to her eating disorder, and therefore be given both diagnoses.)

Criterion E: OCD is rarely due to a GMC/Substance.

Common Pitfall: Newer interviewers may become obsessively concerned with the precise meaning of each word in the definitions and assign a diagnosis to a respondent who does not deserve it. Many people have some obsessive thoughts or compulsive behavior, but OCD is a severe and relatively infrequent condition.

Important Interviewing/Coding Issues: Hoarding

- According to the DSM-IV, if hoarding symptoms are extreme (e.g., if symptoms are accompanied by excessive collecting or buying or stealing of items that are not needed or for which there is no available space), then a diagnosis of OCD is considered. Therefore, severe cases of hoarding discovered in this study will be captured by OCD.

- Symptoms of hoarding include persistent difficulty discarding or parting with possessions, regardless of the value others may attribute to these possessions, due to strong urges to save items and/or distress associated with discarding. These symptoms result in the accumulation of a large number of possessions that fill up and clutter active living areas of the home or workplace to the extent that their intended use is no longer possible (or would if without third-party intervention) and cause clinically significant distress or impairment in social, occupational, or other important areas of functioning (including maintaining a safe environment for self and others).
- Epidemiological studies suggest that hoarding occurs in 2-5% of the population and can lead to substantial distress and disability, as well as serious public health consequences that warrant consideration as a mental disorder. Recent media coverage (e.g., television shows) will likely increase public awareness and self-reported hoarding in this study; however, because we do not ask specifically about hoarding in this study, respondents will not report hoarding unless it is accompanied by some insight (i.e. recognizes that hoarding-related beliefs and behaviors [pertaining to difficulty discarding items, clutter, or excessive acquisition] are problematic)..

G. Generalized Anxiety Disorder (GAD)

Criterion A: This describes the person who may be known by acquaintances as a "worry wart" or a "nervous Nellie." The anxiety or worry is not focused on one or two issues, but is panoramic. "Nellie" worries about the safety of her children, the possibility of being late for an appointment, what to wear to a party, whether her job is in jeopardy, whether there are jellyfish in the water, etc. She worries much of the time, and everyone she knows thinks it is excessive.

Criterion B: Recognizing that the worry is excessive, a person with this problem will often tell him/herself to cut it out, and try to think about something else, but will find him/herself drifting inexorably back to whatever worry is preoccupying him/her at the time.

Criterion F(2): Generalized anxiety is more commonly an associated feature of a Mood or Psychotic Disorder than it is indicative of Generalized Anxiety Disorder. This criterion is presented out of the DSM-IV order in order to allow the interviewer to skip out of the evaluation of GAD if the

anxiety is a feature of another disorder. However, if the period of generalized anxiety clearly precedes the onset of a Mood or Psychotic Disorder, then both diagnoses can be given.

Criterion C: Note that, like the generalized anxiety, some of these symptoms must also be present "more days than not" for a period of at least six months.

Criterion D: Anxiety and worry are important components of many mental disorders. An additional diagnosis of GAD is appropriate only if there are additional symptoms of anxiety and worry that are not part of the other disorder.

Criterion E: This criterion helps to set the boundary between the clinically significant anxiety in GAD and "normal" (e.g., "walking worried"). The anxiety and worry should be considered clinically significant only if they are sufficiently severe to cause marked distress or impairment in functioning.

Criterion F: This criterion instructs the interviewer to consider and rule out a general medical condition or a substance as an etiological factor.

Important Interviewing/Coding Issues:

- Many times a respondent will say "yes" to this at the screener but deny symptoms at the start of the module. Be sure to follow up and ask what he/she was referring to earlier
- If a respondent gives several examples of things he/she worries about, and then says "no" to a 6 month period of time, be sure to follow up and ask how much time he/she does spend worrying. We encourage getting a percentage of time spent worrying regardless of whether the respondent says "yes" or "no" to a 6 month period. You may consider going through the section if the worrying is close to, but less than 50%.
- This 6 month period of worry does NOT need to represent a change for the respondent. Some respondents interpret the question as asking whether there has been a time when he/she worried more than his/her normal level of worry.
- Be sure to assess that the symptoms are present when worrying, and not present when not worrying. Similar to Dysthymia, be aware of chronic fatigue, concentration and sleep problems that may or may not be associated with this anxiety.
- Criterion D (E104) – While it is certainly possible to have both GAD and another Axis I disorder (including another anxiety disorder), use your clinical judgment to assess if one

diagnosis is a better fit than having both – if the symptoms described are better accounted for by another diagnosis.

VII. Module F. Eating Disorders

A. Anorexia Nervosa

Criterion A: You must assess the respondent's lowest weight in the past year, highest weight in the past year, and current weight. It is also necessary to obtain the respondent's height. Although abnormally low weight is necessary for a rating of 3, it is not sufficient. There must be evidence that the person is underweight because of a "refusal" to maintain a normal body weight. The determination as to whether the person's weight is being maintained at a significantly-below-normal level is a clinical judgment. The "below 85%" threshold is merely a guideline and is not intended to be a strict cut-off. Determining "expected" weight ranges can be done by calculating whether the "body mass index" (BMI) [i.e., weight in kilograms/(height in meters)^2] is less than **17.5**. A quick link to a BMI calculator can be found at: http://www.nhlbisupport.com/bmi/. A chart is also included at the end of this Appendix.

Criterion B: This criterion expresses the reason for refusal to maintain a normal weight: an intense (and unreasonable) fear of becoming fat that persists despite the abnormally low weight.

Criterion C: This criterion includes three forms of characteristic distorted thinking, the presence of any one of which would warrant a rating of 3: 1) a marked distortion in the way body size and shape are experienced (e.g., the person is emaciated but still describes a body part that seems "flabby" and needs further reduction); 2) body shape and weight is the central factor in determining self-esteem; and 3) denial of the extent and danger of weight loss.

Criterion D: Irregular periods do not count – there must be a time during which at least three consecutive periods have been missed. Since birth control pills (BCPs) may artificially regulate a menstrual cycle, if someone is having menstrual activity and is on BCP, they are viewed as having met the amenorrhea criterion because the criterion just does not apply. (Same for men and for prepubertal girls.)

B. Bulimia Nervosa

Criterion A: Episodic bursts of binge eating must be distinguished from a pattern of generalized overeating and from isolated episodes of overeating that are context-specific (e.g., at an all-you-can-eat restaurant or a celebration in which there is unlimited food). Binge eating occurs during a discrete period of time, involves consuming a huge number of calories, and is characterized by the sense of having lost control. Although the person may have a feeling of gratification or a reduction in anxiety during the binge, afterwards he/she typically feels uncomfortable, guilty, disgusted, or ashamed. The type of food consumed during binges varies, but usually includes sweet, high calorie treats such as ice cream or cake. Because some respondents may report having had "binges" involving relatively small amounts of food (e.g., eating three cookies), it is important to inquire specifically about the quantity and type of food consumed. Be sure to obtain a concrete example of what the respondent typically eats within a 2-hour period For instance, if a respondent reports eating "a bag of chips," ask about the size of the bag as well A good rule of thumb is that an amount equivalent to two full sized meals is considered enough for a "binge." A respondent must report <u>both</u> a lack of control <u>and</u> a large amount of food consumed.

Criterion B: Binge eating by itself is not sufficient to make the diagnosis. It must be accompanied by inappropriate compensatory mechanisms intended to counteract the effects of binges. The most common of these compensatory behaviors is some form of purging (self-induced vomiting or laxative abuse). Less common compensatory behaviors include fasting, excessive exercise, and manipulation of insulin doses by diabetics. If a respondent reports that he/she is exercising a lot, one might need to probe to determine whether the exercise is in response to a binge, or just a general way of life. Individuals are often very embarrassed about both their binge eating and their compensatory mechanisms (particularly those related to purging). Such information is therefore not volunteered and emerges only with direct questioning.

Criterion C: The minimum frequency of twice a week applies both to the binges and to the compensatory mechanisms, with the presumption that these generally occur together. The binges must be of the level assessed in Criterion A.

Criterion D: This criterion is similar to Criterion C in Anorexia Nervosa and describes the cognitive distortions that are invariably present.

VIII. Module G. Intermittent Explosive Disorder

In order to rate a 3 for this disorder, the respondent must have had several (*at least two*) discrete episodes during the past year in which aggressive impulses were grossly out of proportion to any precipitating factors. These aggressive episodes should not be better accounted for by a Personality Disorder, a Psychotic Disorder, a Mood Episode, or Attention-Deficit/Hyperactivity Disorder.

Important Interviewing/Coding Issues:

- Be sure to get an actual description of a specific recent incident. What was the context? What was the nature of the damages and/or injuries (how severe?) How frequently does this type of thing occur?
- If the behavior is verbal threats, rather than physical violence, the threats must be severe.
- Typically, individuals report a sense of remorse after the fact.
- For the question "Was this on purpose or was it really beyond your control?" reporting that it was beyond the respondent's control would be coded 3.
- Be aware of cultural differences in the acceptance of violence in assessing this disorder.

IX. Module H. Substance Use Disorders

This module contains ratings for the Substance Use Disorders (Dependence and Abuse), which covers problems caused by the respondents' pattern of use of substances. The SCID separates the evaluation of Alcohol from other substances because it is legal, more widely used than other substances, and most users do not have a problem with it.

A. Alcohol Use Disorders

The Alcohol Use Disorders section begins with a series of screening questions to determine whether the respondent's pattern of alcohol use is substantial enough to warrant a detailed evaluation of Alcohol Dependence and Abuse or if the respondent's use of alcohol is insignificant and does not require further assessment. Because respondents often minimize (or underestimate) their drinking habits, do NOT skip the Alcohol section unless the respondent reported drinking fewer than 6 drinks during the past 12 months. Because the definition of Alcohol Abuse requires *recurrent* problems associated with alcohol, infrequent heavy drinking

may still warrant a diagnosis of Alcohol Abuse. If this threshold level of alcohol use is met, establish clear patterns of alcohol use during the past 12 months:

- Typical pattern(s) – the preferred drink (e.g., beer), the amount(s) typically consumed (e.g., 6 12-oz beers), and the frequency of drinking episodes (e.g., 3 nights/wk).
- Heaviest pattern(s) – the same information as above, in addition to the period of time the heaviest pattern occurred during the past 12 months.

Important Interviewing/Coding Issues:

- Ask about the pattern of use again – do not rely on the information obtained earlier in the overview.
- Do not skip the probes on the bottom half of page H1. Information obtained here is useful in assigning the GAF. Get details. If someone objected to their drinking, ask why. Especially ask these questions if the respondent seems to be minimizing his/her use or the effect of his/her use. Imagine a respondent who says he does not have a problem with alcohol, but his spouse is always arguing with him about his use. Why? What does she say?
- The focus on the heaviest period helps prevent the respondent from averaging their usage over the entire year (or minimizing significant events). Remember, however, we are assessing the entire year.
- Questions in the module primarily focus on the period of heaviest drinking. However, if the respondent gives a negative response, also open up the question to the whole year. This is especially important if the period of heaviest drinking was short.
- Use clinical judgment in determining whether a period of time is significant. For example, it is doubtful that a respondent's use would warrant a diagnosis of Dependence for symptoms only occurring during a two week period. However, symptoms that lasted a month or longer may warrant a diagnosis. Use clinical judgment and take the answers to the heaviest period in context.

1. Alcohol Dependence

The criteria have been re-ordered from the DSM-IV in order to make them more user-friendly.

Criterion A(3): The intent of this item is to capture the respondent's failed attempts to put some limits on his/her drinking, e.g., "I'll just have a few beers and then go home" or "I'll stop at the bar for only half an hour." Note that the breaking of these self-imposed limits (e.g., the respondent ends up drinking a couple of six-packs, or staying in the bar for hours) must occur OFTEN in order to be coded 3. There is something of a paradox inherent in the evaluation of this item (and criterion A(4) as well). In order to qualify for these items, the individual must have developed enough insight about having a substance problem to want to control its use. These items are therefore impossible to evaluate in someone who has a very heavy pattern of use but denies any need to control or cut down use.

It is important to note that this symptom is not the same as 'drinking more than one usually does.' The respondent must be putting a limit on him/herself and then going past the limit. Obtain both a description of the amount the respondent planned to consume (or the amount of time he/she planned to use) and how much (or how long) he/she actually drank. As the behavior must occur OFTEN to be coded a "3," it is necessary to assess how frequently this occurred.

Criterion A(4): This item describes unsuccessful attempts to cut down or control drinking over a longer period of time… weeks, months, or years, as opposed to planning for an evening's drinking. Successful attempts to cut down do NOT count towards this symptom. However, successful efforts preceded by several unsuccessful efforts would count. A report that "I may have a problem" is not sufficient to qualify as "persistent desire to cut down." More probes would be necessary.

Criterion A(5): This item covers the ways in which drinking may become a central focus of the respondent's life. As a rule of thumb, two evenings a week is not a "great deal of time"; five evenings a week is, and in between probably justifies a rating of 2.

Criterion A(6): The prototype for this item is the street corner alcoholic who has essentially given up all activities except those associated with drinking. It may also be applied, for example, to an amateur athlete who has stopped doing sports because of drinking, or a person who has stopped seeing all his good friends and now hangs out with a group of heavy drinkers. If a respondent drinks because he doesn't want to socialize (i.e., his alcohol use is not causing him to cut down on important activities; rather he is drinking because he doesn't want to do those activities) he would not receive a code of 3 on this symptom. A respondent would have to have

given up participating in order to drink instead to receive a code of 3. If alcohol is used as a way to avoid an unpleasant activity the symptom would receive a code of 1 or 2.

Criterion A(7): This item is meant to tap a pattern of compulsive alcohol use and does not refer merely to the adverse physical or psychological consequences of drinking. In order to qualify for a rating of 3 on this item, a respondent must understand that the physical or psychological problems are due to the drinking, and still be unable to stop or cut down significantly. Examples of physical problems include cirrhosis or esophageal bleeding due to excessive drinking; examples of psychological problems include alcohol-induced pugnacity that leads to frequent fights and blackouts. The most frequent noxious physical effect of alcohol is a hangover. When hangovers are severe and frequent, and the respondent continues to drink, a coding of 3 is justified on this item.

Criterion A(1): Anyone who drinks at all develops tolerance from the time of adolescent experimenting with alcohol. This item is meant to capture those whose tolerance increased markedly from the time they began drinking fairly regularly to some later time (e.g., "I used to get drunk on 3 beers. Now I can drink 2 six-packs and not be drunk."). It will often be necessary to ask additional probes for this symptom in particular: First, orient the person to the period of time when he/she 'first started drinking regularly' (which is defined as the time when he/she first started drinking at least once per month). Obtain an estimate of how much the respondent needed to feel the effect back then, and during his/her period of heaviest use in the past year. Tolerance is defined as an increase of at least 50%, with a minimum threshold of four drinks. Be aware of respondents who say they have less tolerance now than when they first started drinking. They may be comparing now to a time of heavy drinking, rather than when they first started. Finally, be sure the respondent is answering *not* how much they do drink, but how much it takes to feel the effect.

Criterion A(2): Withdrawal is indicated by the development of the characteristic substance-specific withdrawal syndrome shortly after stopping or decreasing the amount of alcohol. In some cases, the individual never allows the withdrawal syndrome to develop because he/she starts drinking in anticipation of the onset of withdrawal symptoms. Read the list of withdrawal symptoms one at a time.

2. Alcohol Abuse

Criterion A(1): A rating of 3 for this item requires specific evidence that it was the effects of the alcohol use (i.e**.,** either intoxication, withdrawal, or "hangover") that resulted in the respondent's failure to fulfill a major role obligation on at least two occasions. If the level of drinking is significant and the respondent denies any ill effects, explore if others had to cover for the respondent or compensated for him/her.

Criterion A(2): A common error in rating this item is to be over-inclusive and assume that any level of alcohol use in a situation that requires alertness would qualify. This item should be rated a 3 only when the drinking caused sufficient impairment to create a physically hazardous condition (e.g., driving or hunting while intoxicated). Although getting drunk and walking home through a dangerous neighborhood or having unprotected sex with someone one doesn't know very well while intoxicated is certainly risky, neither would warrant a rating of 3—the intent of this item is to rate behavior that puts the respondent in immediate danger because his/her coordination or cognition is impaired by drinking. Recurrent for driving under the influence is 3 or more times.

Criterion A(3): This item should be rated 3 only if legal problems are a direct consequence of the effects of alcohol (e.g., arrest for violent behavior resulting from intoxication). A "Minor in Possession" (MIP) does not count for this symptom.

Criterion A(4): This item is difficult to evaluate when the interpersonal conflict is possibly attributable to a relational problem rather than to a problem with the individual's alcohol use. For example, arguments about occasional non-problematic drinking that are initiated by a spouse who believes any drinking is evil would not warrant a rating of 3.

B. Non-Alcohol Substance Use Disorders

The non-alcohol substance dependence and abuse section is perhaps the most complex part of the SCID because it allows for the simultaneous rating of Dependence and Abuse for 8 different classes of substances. In order to minimize the number of questions that must be asked about each substance, the SCID inquires about the level of drug use for each class to determine whether the Dependence and Abuse questions need to be asked. Read the full drug list to the

respondent and circle the names of all drugs the respondent reports having used in the past year. For all drugs used in the past year, determine the period of heaviest use and the typical and heaviest patterns of use (quantity, frequency) during the past 12 months. For prescription drugs used in the past year, determine whether and how often these drugs have been taken non-medically (i.e., took more of the medication than was prescribed; took the medication for a period longer than was prescribed; used the prescribed medication in a recreational manner [for the feeling or effect only]; used a drug that was not prescribed to him or her [e.g., stolen from a family member, purchased on the street]). Symptoms of Dependence and Abuse are assessed for each drug class for which the threshold level of use has been met. For cannabis, the threshold is at least 6 times in the past year, for prescription drugs, the threshold is nonmedical use at least twice in the past 12 months, and for all other drugs, the threshold is at least twice in the past year. Begin with the drug class meeting the threshold level of use that is highest on the list, query all Dependence and Abuse symptoms for that class, and repeat this process until all other drug classes meeting the threshold level of use have been queried.

Important Interviewing/Coding Issues:

- Most of the points in the Alcohol section (***Section IX.A*** above) apply here as well.
- If the medication is prescribed, be sure to ask about the amount prescribed, and the amount used. Ask whether they ever took the medication for the effect/high.
- Do get pattern of use: typical and heaviest.

1. Non-Alcohol Substance Dependence

Criterion A(3): The intent of this item is to capture the respondent's failed attempts to put some limits on his/her drug use, e.g., "I'll just have one joint tonight." Note that the breaking of these self-imposed limits (e.g., the respondent ends up smoking three joints) must occur OFTEN in order to be coded 3. There is something of a paradox inherent in the evaluation of this item (and criterion A(4) as well). In order to qualify for these items, the individual must have developed enough insight about having a substance problem to want to control its use. These items are therefore impossible to evaluate in someone who has a very heavy pattern of use but denies any need to control or cut down use.

It is important to note that this symptom is not the same as using 'more than one usually does.' The respondent must be putting a limit on him/herself and then going past the limit. Obtain both a description of the amount the respondent planned to consume (or the amount of time he/she planned to use) and how much (or how long) he/she actually used. As the behavior must occur OFTEN to be coded a "3," it is necessary to assess how frequently this occurred.

Criterion A(4): This item describes unsuccessful attempts to cut down or control drug use over a longer period of time…weeks, months, or years, as opposed to planning for an particular evening's decision to use drugs. Successful attempts to cut down do NOT count towards this symptom. However, successful efforts preceded by several unsuccessful efforts would count. A report that "I may have a problem" is not sufficient to qualify as "persistent desire to cut down." More probes would be necessary.

Criterion A(5): This item covers the ways in which drug use may become a central focus of the respondent's life. It is especially variable across classes of drugs because of differences in cost, availability, legality, and the typical pattern of use of the particular substance. For example, the high cost, daily need, and relative unavailability of opioids is much more likely to result in an individual becoming totally preoccupied with the daily task of procuring them. In contrast, this item is less likely to apply to inhalants because of their low cost, wide availability in stores, and the typical pattern of intermittent use. As a rule of thumb, two evenings a week spent smoking pot is not a "great deal of time"; five evenings a week is, and in between probably justifies a rating of 2.

Criterion A(6): The prototype for this item is a heroin addict who has essentially given up all activities except those associated with procuring and using heroin. However, it may also be applied, for example, to an amateur athlete who has stopped doing sports because of using drugs, or a person who has stopped seeing all his good friends and now hangs out with a group of heavy drug users. If a respondent uses because he doesn't want to socialize (i.e., his substance use is not causing him to cut down on important activities; rather he is using drugs because he doesn't want to do those activities) he would not receive a code of 3 on this symptom. A respondent would have to have given up participating in order to use drugs instead to receive a code of 3. If drugs are used as a way to avoid an unpleasant activity the symptom would receive a code of 1 or 2.

Criterion A(7): This item is meant to tap a pattern of compulsive use of the substance and does not refer merely to the adverse physical or psychological consequences of using that substance. In order to qualify for a rating of 3 on this item, a respondent must understand that the physical or psychological problems are due to the substance use, and still be unable to stop or cut down significantly. Examples of physical problems include serious damage to the nasal mucosa from sniffing cocaine or exacerbation of asthma during smoking excessive amounts of marijuana. Examples of psychological problems include cocaine-induced paranoia, or panic attacks precipitated by marijuana.

Criterion A(1): The development of tolerance occurs most frequently with amphetamine, cocaine, nicotine, opioids, and sedatives (especially barbiturates). Tolerance for many drugs (e.g., cocaine, barbiturates, heroin) is usually apparent to the respondent. For drugs like marijuana, where the quality of the drug varies markedly, it may not be possible to establish tolerance. It will often be necessary to ask additional probes for this symptom in particular: First, orient the person to the period of time when he/she 'first started using the substance regularly' (which is defined as the time when he/she first started using at least once per month). Obtain an estimate of how much the respondent needed to feel the effect back then, and during his/her period of heaviest use in the past year. Tolerance is defined as an increase of at least 50%. Be aware of respondents who say they have less tolerance now than when they first started. They may be comparing now to a time of heavy use, rather than when they first started. Finally, be sure the respondent is answering *not* how much they do use, but how much it takes to feel the effect.

Criterion A(2): Withdrawal is indicated by the development of the characteristic substance-specific withdrawal syndrome shortly after stopping or decreasing the amount of the drug. In some cases, the individual never allows the withdrawal syndrome to develop because he/she starts using in anticipation of the onset of withdrawal symptoms. The severity and clinical significance of the withdrawal syndrome varies by class of substance. Characteristic withdrawal syndromes are most apparent with sedatives and opioids. Criteria sets are also provided for withdrawal from amphetamine and cocaine. Although withdrawal symptoms sometimes occur, no specific criteria sets are provided for withdrawal from cannabis, hallucinogens, inhalants, or PCP.

2. Non-Alcohol Substance Abuse

Criterion A(1): A rating of 3 for this item requires specific evidence that the effects of the substance use resulted in the respondent's failure to fulfill a major role obligation on at least two occasions. If the level of use is significant and the respondent denies any ill effects, explore if others had to cover for the respondent or compensated for him/her.

Criterion A(2): A common error in rating this item is to be over-inclusive and assume that any level of substance use in a situation that requires alertness would qualify. This item should be rated a 3 only when the substance use caused sufficient impairment to create a physically hazardous condition (e.g., driving or hunting while intoxicated). Although getting stoned and walking home through a dangerous neighborhood or having unprotected sex with someone one doesn't know very well while intoxicated is certainly risky, neither would warrant a rating of 3—the intent of this item is to rate behavior that puts the respondent in immediate danger because his/her coordination or cognition is impaired by substance use. Recurrent for driving under the influence is 3 or more times.

Criterion A(3): This item should be rated 3 only if legal problems are a direct consequence of the effects of the substance (e.g., arrest for violent behavior resulting from intoxication). Legal problems resulting from procurement or possession of illicit substances do not count because these are so much a function of local laws, attitudes, and enforcement policies.

Criterion A(4): This item is difficult to evaluate when the interpersonal conflict is possibly attributable to a relational problem rather than to a problem with the individual's substance use. For example, arguments about occasional non-problematic substance use that are initiated by a spouse who believes that even minimal drug use is evil and depraved would not warrant a rating of 3.

X. Adjustment Disorder

In most cases, this module is skipped during the administration of the SCID because another more specific diagnosis has already been made. Consider this module only if there is a past year problem described in the overview, but no Axis I disorder has been identified by the SCID to account for it. The border between Adjustment Disorder and ordinary problems of life

may be clarified by the notion that Adjustment Disorder implies that the severity of the disturbance is sufficient to justify clinical attention or treatment.

Important Interviewing/Coding Issues:

- A diagnosis of Adjustment Disorder can be assigned even if another past year diagnosis is present.
- There must be a precipitating event for this disorder to be diagnosed.
- This is the one section that can be coded without actually asking the questions – however it is always better to get more information and confirm.
- It is possible to have on-going stressors – thus the symptoms would persist for more than 6 months. The key here is whether the symptoms persist for more than 6 months *after the stressor or its consequences cease*. The prototypical example of this type of event is post-divorce.

XI. End of Interview

The Global Assessment of Functioning (GAF) scale is contained in this section of the SCID. The GAF scale is designed to measure psychological, social, and occupational functioning. For this study, we will be rating the GAF based on the respondent's worst point during the past year. A score of 100 is assigned to a hypothetical healthiest patient and a score of 1 is assigned to the most ill patient. GAF scores should reflect two components: symptom severity and functioning. Ratings are based on the component that reflects the greatest level of impairment during the past year, rather than an average of the two components. Most respondents will receive a score somewhere between 100 and 1.

There are 4 steps to be followed in determining the appropriate GAF score for a respondent.

Step 1: Starting at the top of the scale, look at each decile and ask yourself, "is either the respondent's symptom severity or functioning worse than what is specified in this range?"

Step 2: If the answer to that question is yes, move down to each subsequent decile until the description for that range matches the respondent's level of symptom severity or functioning (whichever is worse).

Step 3: Double check that the appropriate range has been selected by looking at the range below the one selected and determining that the description for that range is more severe in terms of both symptom severity and functioning than is true for the respondent.

Step 4: Determine the specific number within the 10-point range that is most appropriate for the respondent. This is done by thinking of a hypothetical continuum of persons falling within the chosen range and deciding where the respondent falls on that continuum.

Review your notes about the respondent and ask any additional questions necessary to enable a confident assignment of a GAF score before reading the end of interview script to the respondent and ending the interview. This last point cannot be stressed enough.

XII. Interviewer Debriefing

Complete the questions in this section within an hour of having completed the interview. The answers to these questions provide information about threats to the validity of the data gathered during the interview as well as documentation about unusual situations that may have occurred during the interview session (e.g., the respondent became distressed or showed signs of cognitive impairment, or the respondent was frequently distracted throughout the interview session). This information will be used to determine whether the interview data should be included in the data set to be analyzed or whether the data should be discarded.

Short Blessed Test (SBT) Administration and Scoring Guidelines

A spontaneous self-correction is allowed for all responses without counting as an error.

SB-1. "What year is it now?"

Acceptable Response: The exact year must be given. An incomplete but correct numerical response is acceptable (e.g., 01 for 2001).

SB-2. "What month is it now?"

Acceptable Response: The exact month must be given. A correct numerical answer is acceptable (e.g., 12 for December).

--- "Please repeat this phrase after me: John Brown, 42 Market Street, Chicago."

It is important to carefully read the phrase and give emphasis to each item of the phrase. There should be a one second delay between individual items. The trial phrase should be re-administered until the respondent is able to repeat the entire phrase without assistance or until a maximum of three attempts. If the respondent is unable to learn the phrase after three attempts, a "C" should be recorded. This indicates the respondent could not learn the phrase in three tries.

Whether or not the trial phrase is learned, say "Good, now remember that name and address for a few minutes."

SB-3. "About what time is it?"

This is scored as correct if the time given is within plus or minus one hour. If the respondent's response is vague (e.g., "almost 1 o'clock), they should be prompted to give a more specific response.

SB-4. "Please count backwards from 20 to 1."

The instructions should be read as written. If the respondent skips a number after 20, an error should be recorded. If the respondent starts counting forward during the task or forgets the task, the instructions should be repeated and one error should be recorded. The maximum number of errors is two.

SB-5. "Please say the months of the year in reverse order."

The instructions should be read as written. To get the respondent started, the interviewer may state "Start with the last month of the year. The last month of the year is_____." If the respondent cannot recall the last month of the year, the interviewer may prompt this test with "December"; however, one error should be recorded. If the respondent skips a month, an error should be recorded. If the respondent starts saying the months forward upon initiation of the task, the instructions should be repeated and no error recorded. If the respondent starts saying the months forward during the task or forgets the task, the instructions should be repeated and one error recorded. The maximum number of errors is two.

SB-6. "Please repeat the phrase I asked you to repeat before."

The respondent should state each item verbatim. The address number must be exact (i.e. "4200" would be considered an error for "42"). For the name of the street (i.e. Market Street), the thoroughfare term is not required to be given (ie. Leaving off "drive" or "street") or to be correct (ie. Substituting "boulevard" or lane") for the item to be scored correct.

These guidelines are based on the administration experience of faculty and staff of the Memory and Aging Project, Alzheimer's Disease Research Center, Washington University School of Medicine, St. Louis (John C. Morris, MD, Director & PI; morrisj@abraxas.wustl.edu). They have been modified.

Body Mass Index Chart

BMI	Height (in)																		
	58	**59**	**60**	**61**	**62**	**63**	**64**	**65**	**66**	**67**	**68**	**69**	**70**	**71**	**72**	**73**	**74**	**75**	**76**
Wgt. (lbs)	4'10"	4'11"	5'0"	5'1"	5'2"	5'3"	5'4"	5'5"	5'6"	5'7"	5'8"	5'9"	5'10"	5'11"	6'0"	6'1'	6'2"	6'3'	6'4'
100	21	20	20	19	18	18	17	17	16	16	15	15	14	14	14	13	13	13	12
105	22	21	21	20	19	19	18	18	17	16	16	16	15	15	14	14	14	13	13
110	23	22	22	21	20	20	19	18	18	17	17	16	16	15	15	15	14	14	13
115	24	23	23	22	21	20	20	19	19	18	18	17	17	16	16	15	15	14	14
120	25	24	23	23	22	21	21	20	19	19	18	18	17	17	16	16	15	15	15
125	26	25	24	24	23	22	22	21	20	20	19	18	18	17	17	17	16	16	15
130	27	26	25	25	24	23	22	22	21	20	20	19	19	18	18	17	17	16	16
135	28	27	26	26	25	24	23	23	22	21	21	20	19	19	18	18	17	17	16
140	29	28	27	27	26	25	24	23	23	22	21	21	20	20	19	19	18	18	17
145	30	29	28	27	27	26	25	24	23	23	22	21	21	20	20	19	19	18	18
150	31	30	29	28	27	27	26	25	24	24	23	22	22	21	20	20	19	19	18
155	32	31	30	29	28	28	27	26	25	24	24	23	22	22	21	20	20	19	19
160	34	32	31	30	29	28	28	27	26	25	24	24	23	22	22	21	21	20	20
165	35	33	32	31	30	29	28	28	27	26	25	24	24	23	22	22	21	21	20
170	36	34	33	32	31	30	29	28	27	27	26	25	24	24	23	22	22	21	21
175	37	35	34	33	32	31	30	29	28	27	27	26	25	24	24	23	23	22	21
180	38	36	35	34	33	32	31	30	29	28	27	27	26	25	24	24	23	23	22
185	39	37	36	35	34	33	32	31	30	29	28	27	27	26	25	24	24	23	23
190	40	38	37	36	35	34	33	32	31	30	29	28	27	27	26	25	24	24	23
195	41	39	38	37	36	35	34	33	32	31	30	29	28	27	27	26	25	24	24
200	42	40	39	38	37	36	34	33	32	31	30	30	29	28	27	26	26	25	24

Addendum 1: Skype User Guide

2011 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study
Clinical Interviewer Skype User Guide

Contract No. HHSS283200800004C

Project No. 0211838

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

June 2011

SKYPE USER GUIDE

To enhance quality, the Mental Health Surveillance Study (MHSS) is changing to use Skype for audio recording interviews. Skype is an Internet phone provider which allows users to make voice calls from a computer via an Internet connection. This guide outlines all procedures specific to the use of Skype on the MHSS and should be used in conjunction with your 2011 MHSS ***Clinical Interviewer Handbook***. For reference, ***Exhibit 1*** summarizes the steps involved for using Skype on the Mental Health Study.

A. Equipment

To conduct and record interviews through Skype, use the provided Skype headset. The cushioned headset expands/contracts to fit comfortably. The position of the microphone also adjusts.

The headset attaches to the computer through any of the USB ports on either the left or right side of the laptop and can be connected at any time prior to use. The blue symbol on the left earpiece glows blue when connected.

Use the volume control buttons on the left earpiece (the side with the microphone) and press either the + button at the top or the – button at the bottom to adjust the volume.

B. Skype Account

Placing calls to respondents and recording interviews can be done using a Skype account you will create for your MHSS work. Please note this account should only be used for MHSS work-related calls and should be closed upon the termination of your work on the project. If you have an existing Skype account, you must create another account that is reserved for MHSS work.

B.1 Account Set-Up

While connected to the Internet, first attach your Skype headset then use the Skype link on the desktop to open Skype. Follow the prompts to establish a Skype account, keeping the following in mind:

- Use your MHSS work email address as the basis of your MHSS Skype user name (i.e. clinical.interviewer038). If your user name has already been used, add a leading zero (i.e. clinical.interviewer**0**038).
- Pay attention to the check boxes on the first few introductory screens:
  - Uncheck the "Yes, send me Skype news and promotions" box on the initial 'Create a new Skype account' screen
  - Leave the check box checked for "Sign me in when Skype starts" on the 'Set up your Skype profile' screen
  - Uncheck the "Show welcome screen at start-up" box on the 'Welcome' screen
- Under the Account menu option, click 'See subscriptions' in the middle of the screen

- Sign up for the monthly Unlimited USA and Canada subscription. The charge should be 2.99 per month, charged in 3 month blocks. (Do not select the 12 month payment option.)

- Charge the Skype subscription fees to your personal credit card, then submit the expense for reimbursement on your next ePTE (see ***Section E*** for details).

Note: The 'unlimited' subscription does have limits which should not interfere with your work (maximum of 6 hours per day, 10,000 minutes per month (over 166 hours), no more than 50 numbers called per day). If you expect to reach a limit during an interview, use the telephone recording system. Note calls to other Skype users or to toll free numbers (800, 877, etc.) do not count towards your limits.

OPTIONAL: You could use Skype to call RTI managers/supervisors if you wish. Call the RTI 800 number, and when the recorded voice answers, click the 'Dial pad' on the call menu and enter '1'.

As the voice asks for the extension, enter the 5 digit extension for the person you are calling. Note the '1' will still display on the dial pad but does not affect the call routing. Refer to ***Exhibit 1.1*** in your ***CI Handbook*** for contact information.

Once your account is set up, make several test calls to ensure Skype is working properly. Skype also provides a free 'Sound Test Service' to test Skype.

B.2 Account Settings

Use the following instructions to update the settings of your account.

1. On the Main Skype page, click the **Skype** menu (found on the row with Skype, Contacts, Conversation, etc), then click **Privacy**.
2. Change the main privacy setting to "Only allow people in my Contact list to contact me," as shown.
3. Click "Show Advanced Options."
4. Alter your Privacy settings to match the image below. **Be sure to select "no history"** from the drop down list on the right side of the screen. Click Save when finished.

Following these instructions carefully helps **preserve privacy for you and your respondents**.

5. On the Main Skype page, click the **Call** menu (found on the row with Skype, Contacts, Conversation, etc), then click **Audio Settings**.
6. Check that your settings match those shown in the image to the right.
7. Click "Show Advanced Options."

8. Alter your Audio settings to match the image below. Be sure the Plantronics option is listed first for both the microphone and the speakers. Click Save when finished.

B.3 Caller ID

By calling through an Internet phone provider, your caller ID may appear in a number of different formats, such as "unknown," "0000123456," or "external call unknown number." Skype does allow changes to the caller ID display if you use a cell phone number or a Google voice number, but not if you use a land line. Also, there is no guarantee what number actually might appear.

Understandably, respondents may hesitate to answer a call with a vague caller ID. If the respondent doesn't answer your Skype call, consider using your usual telephone (home/office/cell) and the SAMHSA calling card.

- If the respondent can complete the interview at that time, you must be ready with your equipment and supplies. Tell the respondent you will call right back, stating the caller ID may be different. You can then call again using Skype in order to record the interview.
- If you are setting an appointment, inform the respondent when you call again for the appointment, the caller ID may be different.

Contact your Data Collection Manager with any questions about caller ID.

B.4 Calling with Skype

To dial a respondent's number, use the "Call phones" option at the bottom of the Skype page. **DO NOT enter a respondent number as a contact in Skype**. The respondent's name and number are confidential information to be protected. Dial the number using the 'keypad' on the screen, or to enhance speed and accuracy, copy and paste the number from the CMS or MH Case Manager. Highlight the number, press Ctrl c, then click the "Enter number" space at the top of the Skype keypad, and press Ctrl v. The green phone button at the bottom of the keypad illuminates bright green; click to dial the number.

Be sure to **delete the number from the keypad when finished**. With a number entered, the gray shape to the right of the number becomes an X. Click the X to delete the number.

The setting for "No History" as explained in ***Section B.2*** should remove the records of any prior calls made. If needed, use the "Clear history" button shown on the Privacy Settings screen.

You may choose to use your Skype account for other work related calls, such as to a Clinical Supervisor or your Data Collection Manager. Feel free to enter their names and numbers as Contacts within Skype.

Be sure the volume setting through Skype is maximized for recording purposes. While completing a test call, use the volume/speaker icon (shown in the image below to the right of the mute microphone icon)

to ensure the Skype volume is all the way up. You may adjust the volume to a comfortable level for the headset using the volume up/down buttons on the left earpiece of the headset.

B.5 Three Way Calling with Skype

Although rare, you may need to participate in a three-way conference call through Skype with the respondent and an emergency care representative. Recall MHSS has a very detailed Distressed Respondent Protocol described in the separate Distressed Respondent Protocol Booklet. Scenario 2, with its elevated risk of harm to self or suicide, asks the respondent to remain on the line while you contact emergency care services in the respondent's area.

Skype does allow multiple parties to participate in a call using its conference call feature. Unlike telephone three-way calling, you cannot put the respondent on hold to converse privately with the emergency care representative before linking the respondent and the representative.

To conference call through Skype, follow these steps:

1. Mute the microphone so the respondent cannot hear you (after following the scenario script and explaining your plans): Wiggle the mouse to display the tool bar at the bottom of the Skype screen, then click the microphone icon (to the right of 'End call') to mute the call.

2. Call the appropriate emergency care representative using your usual telephone (home/office/cell) and the SAMHSA calling card. Follow the scenario script to explain the situation to the representative. Obtain the name of the person and explain you will hang up and call again.
3. Reconnect the respondent's call by clicking the mute microphone icon again to unmute.
4. Click "Add people" and call the same emergency care representative.
5. Stay on the line while the representative and the respondent talk to maintain their connection.

Be sure you have another phone available when interviewing in Skype. Although this situation rarely occurs, you must be prepared and able to handle it properly should it happen.

B.6 Ending Skype

When finished with Skype, be sure to Sign Out (through the Skype menu on the main page) or Quit (right click the Skype icon in the system tray). Selecting Close or clicking the red X only closes the window; it does not end your Skype session. Be sure to Sign Out to clear your call history and end Skype.

C. MH Case Manager

The MH Case Manager allows you to record interviews using either Skype or the telephone and a recording control device, with the preferred method being Skype. The first time you open the program with Skype already open, the Case Manager prompts you to allow access to Skype.

Click OK.

Then, within Skype, click the "Allow Access" button in the yellow space at the top of the screen.

C.1 Configure MH Case Manager

To select your recording method, click "Configure" from the menu. From the drop down list, highlight "Recording Mode" then select either "Microphone Line In" or "Skype." Note Skype is the initial setting, so you only need to change it to record a telephone interview. Be sure to change it back to Skype when finished.

An icon in the lower right corner of the MH Case Manager allows you to check the configuration at a glance. For Skype, the icon displays as a blue S, while the telephone recording icon displays as a green microphone. If the Skype program is completely shut down or unavailable on your laptop, the Skype S shows as gray instead of blue.

C.2 Recording the Interview with Skype

While connected to the Internet, have the MH Case Manager open and the case selected. Start Skype as needed (the Case Manager should already start Skype for you). You can arrange the two windows on your laptop screen to easily view and use both programs.

Complete your test call by first clicking "Recording Test" and then dialing the number through Skype. Note the "Start Recording" button in the Recording Test box **does not turn green** until the actual connection is made. So, if you call your cell phone voice mail as your test call, once the connection is made and your voice mail message starts to play, be ready to click "Start Recording." You will miss recording the first few seconds of the message, but that is not a problem.

When you call a respondent, again, the "Start Recording" button will be grayed out and not turn green until the connection is made. Since you wait to begin recording until the respondent provides permission during the introduction process, the slight delay is not a problem.

When finished recording, click "Stop Recording." Take care to not accidentally click too early.

C.3 Uploading Files

Uploading completed recordings remains the same straightforward process. While connected to the Internet, click "Upload." There is one MH Case Manager difference to note: At the bottom of the screen, Skype files display a "Mixed" check box along with a "Mixed Date/Time." These items will be blank for telephone recordings.

File Name		WAV Size (KB)	MP3 Size (KB)	Record Start Date/Time	Record Stop Date/Time	Mixed?	Mix Date/Time	Converted?	Convert Date/Time	Uploaded?	Up Da
000999_9999999_20110601114100		369.004	26.28	06-01-2011 11:41 AM	06-01-2011 11:41 AM	☑	06-01-2011 11:41 AM	☑	06-01-2011 11:41 AM	☑	06
000999_9999999_20110601114147		297.324	20.484	06-01-2011 11:41 AM	06-01-2011 11:41 AM	☑	06-01-2011 11:41 AM	☑	06-01-2011 11:41 AM	☑	06
000999_9999999_20110601131109		409.004	27.468	06-01-2011 1:11 PM	06-01-2011 1:11 PM	☑	06-01-2011 1:11 PM	☑	06-01-2011 1:11 PM	☑	06
000999_9999999_20110601150042		262.444	18.036	06-01-2011 3:00 PM	06-01-2011 3:00 PM	☑	06-01-2011 3:00 PM	☑	06-01-2011 3:00 PM	☑	06
000999_9999999_20110601153105		209.382	64.728	06-01-2011 3:31 PM	06-01-2011 3:31 PM	☐		☑	06-01-2011 3:31 PM	☑	06

Additionally, a blue Skype icon or a green microphone icon displays in the 3rd column indicating the recording method for the file. The case information bar for uploaded, unselected cases displays white.

Exhibit 1 MHSS Skype Call Summary Steps

Set Up Tasks	Reference
Establish Skype account	Section B.1
Update Skype privacy settings	Section B.2
Configure MH Case Manager for Skype	Section C.1

Respondent Call Tasks	Reference
Prepare materials	CI Handbook, Chapter 4
Prepare equipment - Laptop connected to Internet - Skype headset attached - MH Case Manager open and case selected - Skype open	Section A
Complete test recording	Section C.2
Call respondent using 'regular' calling method (if can do interview now, call back with Skype)	Section B.3
For an interview appointment, call using Skype	Section B.4
Record interview	Section C.2
Delete number from Skype keypad when finished	Section B.4
Upload file	Section C.3
Complete all other documentation steps	CI Handbook, Chapter 6

D. Trouble Shooting

In testing, Skype has proven reliable and user-friendly while greatly enhancing the quality of the interview audio recordings. However, problems may arise. Keep the following in mind:

- **Monitoring Call Quality**: While there are no specific steps necessary for you to monitor call or recording quality, please contact your Data Collection Manager if you encounter any problems. If there are any issues with the quality of your submitted recordings, a supervisor will contact you.
- **Avoiding Interrupted Recordings**: A particular laptop setting disables the 'Tap' feature on the Touchpad. This avoids the inadvertent brush of the Touchpad that generates a tap/click on the "Stop Recording" button during an ongoing interview. Please do not change this setting.
- **Avoiding Dropped Calls**: We hope this will occur rarely, but dropped calls may happen. Sometimes it depends on the quality of the Internet connection. If your wireless has occasional breaks in service, perhaps try an Ethernet connection which can be more stable.
- **Handling Dropped Calls**: Should you experience a dropped call, try to call the respondent again quickly, using a telephone as needed. Explain you had an equipment issue; you'll try to resolve it quickly and then will call them back. If your Internet returns quickly, call back through Skype to complete the interview. If the issue cannot be resolved quickly (maybe a storm induced power outage) call the respondent to set an appointment.
- **Avoiding Skype Overload**: When using Skype, minimize the number of open applications on your laptop. Clearly you must have the MH Case Manager open along with Skype, and it can be helpful to have the CMS open as well. If you encounter a distressed respondent situation and need to use the SAMHSA website and/or the National 911 Registry, those applications can also be used in conjunction with Skype. Remember the laptop should only be used for MHSS work-related tasks.
- **Conducting an Interview without Skype**: As needed, use the telephone audio recording method described in ***Chapter 3*** of your ***CI Handbook*** to record an interview. Change the MH Case Manager configuration to Microphone Line In (see ***Section C.1***). Handbook page 3-1 describes the use of separate headphones to disable the laptop speakers and allow you to listen to the test recording. Do NOT use separate headphones, and do NOT use the Skype headset. Instead, mute the laptop speakers following a successful test recording call. When finished interviewing, unmute the speakers and change the configuration back to Skype.
- **Conducting an Interview without Internet**: If you need to conduct an interview without any Internet access, refer to the above bullet and also page 3-10 of your ***CI Handbook***.

Should you have any issues using Skype, please contact your Data Collection Manager. The managers need to know what issues you are facing and are glad to provide assistance. They may direct you to contact the project's Technical Support Group (TSG) at .

E. Administrative

Please note the following information for handling administrative details related to Skype.

E.1 Subscription Reimbursement

As mentioned, you will charge the 3 month Skype subscription fee to your personal credit card. To be reimbursed, record the expense on your next ePTE. You don't have to wait for the charge to appear on your credit card bill, since you know the total.

- Record the total amount in the expense section on the back of your CI ePTE Working Copy.
- Divide the expense into the 1/3 and 2/3 split. Refer to ***Chapter 7*** of your ***CI Handbook*** as needed.
- Keep in mind expenses over $10 must be submitted on a Headway Expense Report. Given the current Skype rate for a 3 month subscription, this should not be an issue.

Some credit cards may charge a small fee (0.30 or so) for processing an international payment (recent Skype charges were processed through a Luxembourg location). Such a fee would be a reimbursable expense.

E.2 Subscription Renewal

As the end of your 3 month Skype subscription nears, you will be notified by Skype. Unless otherwise directed, renew your subscription for an additional 3 months. Charge a personal credit card, and then claim the expense on an ePTE as explained above.

Remember this account should only be used for MHSS work-related calls, and should be closed upon the termination of your work on the project.

As always, if you have questions about or issues with using Skype, contact your Data Collection Manager.

This page intentionally left blank.

Addendum 2: Distressed Respondent Protocol Booklet

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study
Distressed Respondent Protocol Booklet

Contract No. HHSS283201000003C

Project No. 0212800

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

March 2012

Quick Reference Guide

Distressed Respondent Situation Contacts:

Call:	Phone Number:

National Lifeline Number:

Locating a Respondent's Address:

1. Navigate to the CMS "Case Status" link
2. Locate the QuestID/Case Data information for the case
3. Click "Click Here" in the Address column on the far right

SAMHSA Mental Health Services Referral Information:

http://mentalhealth.samhsa.gov/databases/
(link on your laptop's Desktop)

Using the National 911 Registry (see page DRP-10 for details):

1. Use the County Search desktop link: http://www.naco.org/Counties/Pages/CitySearch.aspx and enter either the respondent's City or ZIP code to find the county
2. Use the 911 Registry file from either the desktop link or the CMS
3. Use the corresponding Region tab to locate the state and county
4. Locate the 24 hour number, selecting the most appropriate number for the R's address

3 Way Calling:

SKYPE

1. Mute microphone for respondent call
2. Using another telephone, contact the emergency care representative, explain the situation, obtain their name, and indicate you will call again. End the call
3. Unmute the respondent call
4. Through Skype, add the emergency care representative to the call (Add people)
5. Stay on the line to maintain the connection between the respondent and the representative

OWN PHONE (record the steps for your phone here)

1.
2.
3.

MHSS DISTRESSED RESPONDENT PROTOCOL

Distressed Respondents

Due to the sensitive nature of portions of the SCID interview, you may encounter a respondent who becomes upset during the interview. In situations where you feel that based on indirect and direct statements made by the respondent that he or she has had suicidal or homicidal thoughts within the past two weeks or is otherwise distressed, use the Distressed Respondent Protocol (see ***Exhibit A***) to determine the proper course to follow. So that it is always readily available, this Protocol is printed at the back of each SCID booklet. The first page of the protocol lists issues to be aware of before, during, and following the interview as you interact with the respondent. Although you may be qualified, you are NOT to provide counseling or other services. For your work on this study, you must follow the approved protocol.

The chart at the bottom of the first page lists various scenarios you may encounter and refers to the more detailed instructions on the following pages. Deciding which scenario to use is a judgment call for you to make based on your interactions with the respondent.

In certain scenarios and in ANY case where you are unsure of the proper scenario or correct approach, you are to consult with project staff to determine the best course of action for the situation. The Clinical Supervisors ( ), who are licensed psychologists, will act primarily as a sounding board for you. If there is a question about what action to take in response to your interactions with a respondent, you will discuss the situation with one of the Clinical Supervisors, and the Clinical Supervisor will make the final decision as to what action, if any, should be taken beyond documenting the situation. These steps are for your protection and are required by RTI's IRB.

In accordance with the Distressed Respondent Protocol or after consulting with one of the Clinical Supervisors ( ), you may need to call an emergency psychiatric care provider (e.g., crisis center, inpatient hospital) or 911 in the respondent's area so emergency help can be dispatched to the respondent's location as needed. Obtain referral information for the respondent's area from the SAMHSA website (http://mentalhealth.samhsa.gov/databases/) and the national 911 database (see ***Exhibit B***).

When calling an emergency psychiatric care provider, use 3-way calling to keep the respondent on the line while you call. Be familiar with and practice using 3-way calling prior to initiating any interviews so you can perform this operation smoothly, if needed. Note: with 3-way calling, once the respondent is speaking with the other person, you must **remain on the line** in order to maintain their connection.

Once speaking with an emergency psychiatric care representative or 911 in the respondent's area, use the script in the Distressed Respondent Protocol to explain the situation, and provide the respondent's name, telephone number, and, if needed, the address reported by the respondent (if known) as well as the respondent's home address (if the respondent reports a different location). To protect respondent confidentiality, when possible, only share information required to properly assist the respondent. The emergency care representative or 911 dispatcher will need the entire address to send assistance.

To report the respondent's address to the emergency psychiatric care representative or 911 dispatcher:

- ask the respondent for the address of his/her current location; and
- obtain the respondent's home address from the CMS. From the CMS, access the case (through the Case Status link), and click as directed in the Address column to display the address for the QuestID. To protect confidentiality the address is not provided as assignment information, but is readily available should you need it in an urgent situation.

While you are not to participate in the conversation between the respondent and the emergency care representative, do pay attention to be sure the respondent receives instructions for how to receive emergency services. If the emergency care representative is not able to provide services to protect the safety of the respondent, call 911 in the respondent's area. Note the protocol includes these instructions to ensure proper handling of each respondent's situation. If an emergency care representative asks you to stay on the line with the respondent while the rescue team arrives, do not complete the interview. Rather, keep the conversation casual and light, avoiding topics that may be upsetting to the respondent.

When finished with any distressed respondent situation, be sure to document both the situation and your response by completing an online Incident Form accessible through the Main Page of the CMS. See ***Exhibit C*** for the types of incident information to enter. **Proper documentation of all distressed respondent situations is critically important** to protect respondents' rights. For non-emergency situations (Scenarios 1, 3, 5), contact to consult with them about the course of action that may need to be taken to protect the respondent or other possible victims. If you have questions about what to include in an Incident Report, or if you want to debrief about the case, you may also call . Additionally, you must contact immediately for any cases for which imminent danger is or may possibly be involved (Scenarios 2 and 4) to receive supervision and debrief. If it is an urgent situation and is unavailable within 15 minutes, try the next person on the list, and so on. In addition to the study contact summary printed in the inside cover of this handbook, the chart below lists these contacts:

Call:	Cell Phone Numbers:

Please note that do not have clinical training. Calling one of them is important to alert project staff that the Distressed Respondent Protocol has been enacted and you are attempting to contact the Clinical Supervisors ( ).

The study's IRB approval rests on following these carefully developed procedures exactly.

Exhibit A Distressed Respondent Protocol

Specific Guidelines

If respondents report any of the issues listed below during any interactions with the recruiter or clinical interviewer, including before, during, or after a telephone screening or interview, the staff member will immediately refer to the scenario chart below and follow the instructions provided. Details of all incidents will be documented on the case management system and reported to project management staff immediately.

- Has had **any suicidal thoughts in the past two weeks (p. A.3)**, including
  - passive suicidal thoughts (i.e. thoughts or wishes about his/her death **in the absence of** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, or intention of dying or attempting suicide) **[SCENARIO 1]** or
  - active suicidal thoughts (i.e. thoughts or wishes about his/her death **combined with** thoughts about specific ways s/he could die or attempt suicide, plans for how s/he could die or attempt suicide, the intention of dying or attempting suicide, the means to carry out that plan, and will not promise you that s/he will not hurt him/herself) **[SCENARIO 2]**

- Has had **any homicidal thoughts in the past two weeks**, including
  - passive homicidal thoughts (i.e. thoughts or wishes about seriously harming someone else **in the absence of** thoughts about specific ways in which s/he could seriously harm another person, plans for how s/he could seriously harm another person, intentions of seriously harming another person) **[SCENARIO 3]** or
  - active homicidal thoughts (i.e. thoughts or wishes about seriously harming someone else **combined with** thoughts about specific ways s/he could seriously harm another person, plans for how s/he could seriously harm another person, the intention of seriously harming another person, and the means to carry out that plan) **[SCENARIO 4]**

Scenario Chart		
Scenario Number	**Individual at Risk of Harm**	**Imminent Danger?**
1	Self	No
2	Self	Possible / Yes
3	Other(s)	No
4	Other(s)	Possible / Yes
5	No risk of harm; respondent is agitated or upset	No

Exhibit A Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
1	Self	No

STEPS

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about your death or dying.

B. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about death and dying. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

C. WHEN CALL IS COMPLETED, CALL IF YOU HAVE QUESTIONS OR WOULD LIKE TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

Exhibit A Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
2	Self	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you are thinking about harming yourself. Can you promise me that you will not harm yourself today?

IF NO: GO TO STEP B.

IF YES: Do you have a doctor, counselor, or other professional you can talk to about how you are feeling?

IF YES: I strongly suggest that you contact this person so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about death and dying. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. I strongly suggest you contact counselors at Lifeline. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. I strongly suggest that we contact emergency care services in your area, such as a crisis center or nearby hospital. I am going to look-up that number. Can you remain on the line while I do that? It may take a few minutes.

IF NO: Okay, if I don't connect you with the local emergency care provider, then I will need to call the provider myself to see if they can send someone to you who can provide the care you need in order to keep you safe. I'll call you back to let you know what I find out.

C. FIND THE NEAREST EMERGENCY PSYCHIATRIC SERVICES USING THE SAMHSA WEBSITE (http://mentalhealth.samhsa.gov/databases/). **SEARCH FOR <u>INPATIENT</u> MH TREATMENT USING THE R'S CURRENT ZIP CODE.**

SCENARIO 2 (Continued)

D. CALL THEIR LOCAL INPATIENT PSYCHIATRIC CARE FACILITY OR CRISIS CENTER AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming (himself/herself) and I am concerned about (his/her) safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

> **IF ASKED FOR NSDUH OVERVIEW:** This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (his/her) well-being, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

E. GIVE R FIRST NAME, TELEPHONE NUMBER, AND ADDRESS (IF KNOWN) TO LOCAL EMERGENCY CARE REPRESENTATIVE. IF THEY ARE UNABLE TO PROVIDE SERVICES THAT ENSURE THE R'S SAFETY, SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

F. IF R NOT ON THE OTHER LINE, END CALL WITH THE EMERGENCY CARE PROVIDER OR LOCAL 911 DISPATCHER AND ATTEMPT TO CONTACT R AGAIN WITH AN UPDATE.

IF R ON THE OTHER LINE, CONNECT R TO EMERGENCY CARE REPRESENTATIVE OR LOCAL 911 DISPATCHER AND STAY ON THE LINE; IF YOU HANG-UP, THEIR CONNECTION WILL ALSO END.

G. YOU MAY STAY ON THE LINE TO WAIT FOR THE RESCUE TEAM TO ARRIVE. IF SO, DO NOT CONTINUE THE INTERVIEW. KEEP THE DISCUSSION LIGHT AND AVOID EMOTIONAL TOPICS. DEMONSTRATE EMPATHIC LISTENING BUT REFRAIN FROM COUNSELING OR PRACTICING PSYCHOLOGY.

H. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Exhibit A Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
3	Other(s)	No

STEPS

A. COMPLETE SCREENING/INTERVIEW AND THEN READ TO R: When you agreed to participate in this interview, I promised that I would not tell anyone what you have told me unless it was necessary to protect you or other people. You told me earlier that you have recently had thoughts or wishes about seriously harming someone else. Do you have a doctor, counselor, or someone you can talk to about how you are feeling now?

IF YES: I strongly suggest that you contact this person immediately so you can talk to him or her about how you have been feeling, especially about the thoughts you've been having about seriously harming someone else. Would you be willing to do that?

IF YES: Okay. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I strongly suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline has counselors available 24-hours a day to talk to you about how you are feeling. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away. If you are not able to get to an emergency room immediately, you should call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF DIRECTED BY ONE OF THEM, FOLLOW SCENARIO 4 FOR POSSIBLE IMMINENT DANGER TO OTHERS. FILL OUT ONLINE INCIDENT REPORT.

Exhibit A Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
4	Other(s)	Possible / Yes

STEPS

A. END SCREENING/INTERVIEW AND END CALL.

B. SEARCH FOR THE R'S LOCAL EMERGENCY NUMBER USING THE NATIONAL 911 DATABASE.

C. CALL THEIR LOCAL 911, AND READ THIS STATEMENT: I work for RTI International, a research company in North Carolina, and we are conducting a research study. During an interview with a respondent, the respondent told me that (he/she) is thinking about killing or harming another individual. I am concerned about this individual's safety. I can give you additional information about the research study, if you would like. I can also provide you with the respondent's contact information.

> **IF ASKED FOR NSDUH OVERVIEW:** This study, part of the National Survey on Drug Use and Health sponsored by the United States Public Health Service, is designed to test procedures for use in future NSDUH surveys. Questions ask about various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence. Please note that this information was obtained through the respondent's participation in a research study. We went through appropriate informed consent procedures, during which I told the respondent that if (he/she) told me something that caused me to be concerned about (him/her) harming someone else, I would report that to someone else who could help or intervene. Given the context in which the information was obtained, however, we cannot guarantee that the participant understood the questions nor that (he/she) provided truthful responses. Do you have any questions about the study? **ANSWER QUESTIONS.**

D. GIVE R FIRST NAME, TELEPHONE NUMBER, ADDRESS (IF KNOWN), AND VICTIM'S IDENTIFYING INFORMATION TO LOCAL 911 DISPATCHER. END CALL.

E. WHEN CALL IS COMPLETED, CALL TO DEBRIEF. IF SHE DOES NOT RETURN CALL WITHIN 15 MINUTES, CALL TO DEBRIEF. IF NEITHER ONE OF THEM IS AVAILABLE, CONTACT TO NOTIFY ONE OF THEM ABOUT THE INCIDENT. FILL OUT ONLINE INCIDENT REPORT.

Exhibit A Distressed Respondent Protocol (continued)

Scenario Number	Individual at Risk of Harm	Imminent Danger?
5	No risk of harm; respondent is agitated or upset	No

STEPS

A. END SCREENING/INTERVIEW AND THEN READ TO R: I know these questions are very personal, and they seem to be upsetting you. Do you have a doctor or someone you can talk to about how you are feeling?

IF YES: I suggest that you call that individual immediately so that she or he can help you talk about and work through how you are feeling. There is also a national Lifeline hotline you can call where counselors are available to talk at any time of the day or night. Their toll-free number is 1-800-273-8255. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

IF NO: I suggest that you contact the national Lifeline hotline at 1-800-273-8255. Lifeline is a 24-hour hotline that you could call to discuss this with a counselor. They may also help you locate (additional) mental health services in your area. If you feel that this is an emergency now or later, you should go to a hospital emergency room right away or call 911 for assistance. **THANK R FOR THEIR PARTICIPATION IN THE STUDY AND END CALL.**

B. WHEN CALL IS COMPLETED, CALL IF YOU HAVE ANY QUESTIONS OR NEED TO DEBRIEF. FILL OUT ONLINE INCIDENT REPORT.

Exhibit B 911 Registry Search Steps

For Scenarios 2 and 4, you may need to contact a local 911 service in the respondent's area. To locate a number for the local area, we have contracted with NENA, the National Emergency Number Association, to provide a copyrighted listing of local service providers within the United States.[2]

Follow these steps to locate a local 911 number:

1. Locate the address of the respondent from the CMS

2. Use the County Search desktop link: http://www.naco.org/Counties/Pages/CitySearch.aspx and enter either the City or ZIP code to find the corresponding county

3. Open the national 911 Registry file (link found on CMS or on your desktop). The file contains the following information:
   - Public Safety Answering Point (PSAP) Name
   - County
   - PSAP Address (Street, Community, State, ZIP)
   - 24 hour Phone Number
   - Second 24 hour Phone Number
   - PSAP Service Area Description

4. Locate the state and the corresponding Region tab to use (the listing below sorts by Region while the INDEX tab in the file shows States in alphabetical order)

Region	States Included
Northeast	Connecticut, Delaware, Massachusetts, Maryland, Maine, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Virginia, Vermont, Washington, DC, West Virginia
Southeast	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas
Central	Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, Wisconsin
West	Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, Wyoming

5. Scroll to the appropriate state and county (alphabetized in the second column)

[2] Copyright 2005-2006 by National Emergency Number Association (NENA) and its providers, all Rights Reserved. Not to be copied, used or distributed in any manner, except in accordance with a license agreement between the user and NENA.

Exhibit B 911 Registry Search Steps (continued)

6. Locate the 24 hour number for 911 services for the county and the respondent's address. Keep in mind the following:
   - Many counties have multiple local 911 numbers: Select the most appropriate number. For example, Anne Arundel County in Maryland has a number for the city of Annapolis and another number for the rest of the county.
   - Densely populated areas may list many individual numbers for the various cities/municipalities within the county. Use the respondent's city to narrow down the choices. If the city is not listed separately, try the county number.
   - The last column in the file contains a description of the service area. This may also help in selecting a number.
   - Some addresses may be physically located in one county, yet the closest 911 provider may be just across the county (or even state) line. Should this happen, the dispatcher answering your call should provide the better/closer number or even be able to transfer your call.
   - Rarely, a PSAP requested that their local number be suppressed in the file so the 24 hour number column appears blank. If this occurs, locate a nearby number. Perhaps there is another listing for the county to try. If not, try to find a nearby county: consider the ZIP code of the address and try to find another listing for that state with a similar ZIP code. If that doesn't work, try another number in the state, requesting assistance in determining the correct number.
   - Do your best to determine the correct number. Even if the number is not exact, you'll have access to a nearby number close to the proper location, and those responders should be able to direct you to the proper number.

Periodically, the 911 Registry file will be updated and posted to the CMS. Your Data Collection Manager will send an email when an updated file is available, with a reminder to copy the file from the CMS and save it to your desktop to ensure you have the most current version.

If you have any questions about accessing or using the 911 Registry, please contact your Data Collection Manager.

Exhibit C Incident Report Form

Incident Report Form

When completing the Incident Report Form, DO NOT include identifying information such as the respondent's name or address

*required fields

Project Title:	National Survey on Drug Use and Health (NSDUH): Mental Health Surveillance Study
***Date of Incident:**	5 / 20 / 2011
***Scenario**	Select
***Respondent's Location:**	
***Participant's QuestID :**	
***Describe in detail the nature and timing of this incident:**	6000 characters remaining
If Report or Referral was made to an outside agency:	
Date and time of report/referral:	MM/DD/YYYY / / HH:MM AM/PM : N/A
Agency and person who took report/referral:	
Current disposition of situation:	1000 characters remaining

Save | Submit Report | Reset | Cancel

Be sure the Incident Report contains essential details for understanding the incident and the actions taken to handle the distressed respondent, including:

- The scenario number that was used,
- The nature and extent of the suicidal thoughts (Scenario 1, Scenario 2), homicidal thoughts (Scenario 3 and Scenario 4), or emotional distress (Scenario 5),
- Information considered in the risk assessment to include any risk factors (e.g., respondent could not contract for safety) and protective factors (e.g., respondent denied a plan, has no history of attempts, respondent is under the care of a psychiatrist),
- What actions were taken to resolve the situation (e.g., referred respondent to Lifeline, connected respondent with the local crisis center, contacted for supervision), and

  Any unforeseen circumstances that arose (e.g., respondent hung-up before script was read.

Appendix E: Mental Health Surveillance Study
Field Interviewer Handbook

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

MHSS
Field Interviewer Handbook

Contract No. HHSS283201000003C

Project No. 0212800

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

October 2011

Field Interviewer:_____ FI ID No:_____

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study Field Interviewer Handbook

Contract No. HHSS283201000003C

Project No. 0212800

Prepared for:

Substance Abuse and Mental
Health Services Administration
Rockville, Maryland 20857

Prepared by:

Research Triangle Institute

October 2011

Research Triangle Institute

MISSION

To improve the human condition by turning knowledge into practice.

VISION

To be the world's leading independent research organization, recognized for solving critical social and scientific problems.

VALUES

Integrity - We perform with the highest ethical standards of individual and group honesty. We communicate openly and realistically with each other and with our clients.

Excellence - We strive to deliver results with exceptional quality and value.

Innovation - We encourage multidisciplinary collaboration, creativity and independent thinking in everything we do.

Respect for the Individual - We treat one another fairly, with dignity and equity. We support each other to develop to our full potential.

Respect for RTI - We recognize that the strength of RTI International lies in our commitment, collectively and individually, to RTI's vision, mission, values, strategies and practices. Our commitment to the Institute is the foundation for all other organizational commitments.

Fiscal Responsibility - We operate with financial integrity and transparency. We are accountable for cost competitiveness and continuing financial responsibility.

Objectivity - Our work is independent of undue influences by political, economic, or other factors. We maintain the highest level of scientific objectivity in our work.

TABLE OF CONTENTS

1. INTRODUCTION TO THE STUDY 1-1

1.1 Overview of the Mental Health Surveillance Study 1-1

1.2 Organization 1-1

1.3 Schedule 1-1

1.4 Sample 1-1

1.5 Use of the Mental Health Study Handbook 1-2

1.6 Field Interviewer Responsibilities 1-2

2. EQUIPMENT AND MATERIALS 2-1

2.1 Introduction 2-1

2.2 Equipment 2-1

2.2.1 Technical Support Procedures 2-1

2.3 Terminology 2-1

2.4 Materials 2-2

2.4.1 Screening 2-2

2.4.2 Initial Interview 2-3

3. INITIAL INTERVIEW 3-1

3.1 Introduction 3-1

3.2 Initial Interview 3-1

3.3 Incentive Procedures 3-1

3.4 Recruitment Process 3-1

3.5 Transmit Each Day You Work 3-10

4. OBTAINING PARTICIPATION 4-1

4.1 Answering Respondent Questions 4-1

4.2 Refusals 4-2

4.2.1 Initial Interview Refusals 4-2

4.2.2 Follow-up Interview Refusals 4-2

4.3 Other Non-response 4-2

4.3.1 Respondent Unavailable 4-2

4.3.2 No Phone Access 4-3

5. ADMINISTRATIVE PROCEDURES 5-1
5.1 Introduction 5-1
5.2 Project Charge Code 5-1
5.3 Completing and Submitting Weekly ePTE 5-1
5.4 PT&E Summary Data 5-1
5.5 Incentive Advances 5-1
5.5.1 Incentive Advance Balance Sheet 5-2
5.6 Transmitting to RTI 5-2

6. QUALITY CONTROL 6-1
6.1 Introduction 6-1
6.2 Data Quality Reminders 6-1
6.3 Verification 6-1
6.4 Ongoing Training 6-1

LIST OF EXHIBITS

Exhibit 2.1 Follow-up Study Description 2-4
Exhibit 2.2 Follow-up Interview Payment Receipt 2-5
Exhibit 2.3 Reminder Card 2-6

Exhibit 3.1 Mental Health Study Initial Interview Practice Cases 3-11

Exhibit 4.1 RECRUIT1 Screen 4-4
Exhibit 4.2 RECRT4WK Screen 4-5
Exhibit 4.3 REFFEAS Screen 4-6
Exhibit 4.4 Respondent Questions and Responses 4-7

Exhibit 5.1 Properly Completed Working Copy ePTE 5-3

This page intentionally left blank.

1. INTRODUCTION TO THE STUDY

The Mental Health Surveillance Study (MHSS) is a special research study being conducted as part of the 2012 National Survey on Drug Use and Health (NSDUH). This handbook outlines all procedures specific to the Mental Health Surveillance Study and should be used in conjunction with your 2012 NSDUH ***FI Manual*** and ***FI Computer Manual***.

1.1 Overview of the Mental Health Surveillance Study

The Mental Health Surveillance Study (also referred to as the Mental Health Study throughout this handbook) was first conducted as part of the 2008 NSDUH. This study is designed to assess the NSDUH's Mental Health module questions by:

- collecting data from all adult interview respondents through a Mental Health module within the initial NSDUH interview; and
- recruiting and completing in-depth follow-up telephone interviews with selected adult respondents.

The follow-up telephone interviews will be completed with respondents aged 18 and older who complete the initial NSDUH interview using an abbreviated version of the Structured Clinical Interview for DSM-IV Axis I Disorders, also known as the SCID. The goal is to complete 1,500 follow-up telephone interviews in 2012.

As field interviewers (FIs) on the MHSS, you are responsible for completing the screening and the initial NSDUH interview. For selected interview respondents, you will use special recruitment scripts displayed in the initial interview to recruit the respondent for the follow-up telephone interview. The follow-up telephone interview will be conducted by a set of specially trained telephone interviewers.

1.2 Organization

The MHSS management structure is the same as the 2012 NSDUH data collection management structure. You will report to your field supervisor (FS) for NSDUH and Mental Health Study work.

1.3 Schedule

The MHSS uses the same quarterly schedule as the 2012 NSDUH. Data collection will take place in all four quarters of 2012.

1.4 Sample

The Mental Health Study is being conducted in all 50 states and Washington, DC. All NSDUH FIs will take part in this special study.

The MHSS sample is embedded within the 2012 NSDUH sample. A sub-sample of English-speaking respondents aged 18 or older who complete the initial NSDUH interview will be selected to complete the follow-up telephone interview in English. Persons aged 12 to 17 years old and respondents who complete the initial NSDUH interview in Spanish are not eligible for the follow-up interview.

The target number of completed follow-up telephone interviews for 2012 is 1,500, or approximately 375 follow-up interviews per quarter. If an interview respondent is selected for the follow-up interview, special recruitment scripts will display during the back-end CAPI portion of the initial interview. Under no circumstances should you mention the possibility of a follow-up interview to the respondent before this time.

1.5 Use of the Mental Health Study Handbook

This handbook provides a detailed description of the additional tasks you are required to complete for this study. It outlines all procedures and materials that are specific to the MHSS, so it should be used in combination with your ***FI Manual*** and ***FI Computer Manual***. Adherence to prescribed procedures and duties is critical to the success of the Mental Health Study. This handbook should be carefully studied as you prepare for fieldwork, and referred to during data collection as questions arise. Keep it in your laptop bag so it is available for reference while working. Take care to not show respondents the handbook; it is for your use only.

You may, however, have questions or encounter field situations for which you do not find an answer in this handbook. When in doubt about any field situation, contact your FS.

1.6 Field Interviewer Responsibilities

With every scientific survey, there are many necessary components that contribute to the overall success of the research. You are responsible for one of the most important aspects of the research: making sure the NSDUH interview is administered properly according to survey procedures.

Specifically, you are responsible for:

- completing the screening and initial NSDUH interview;
- following the recruitment scripts to recruit selected respondents for the follow-up telephone interview;
- transmitting each day that you work (Your timely transmission is crucial due to the tight schedule for completion of the follow-up interview.); and
- adhering to the current NSDUH data collection standards you are already well acquainted with, as well as the procedures described in this handbook.

It is imperative that you follow all procedures and protocols as you are working the MHSS. You have already signed a NSDUH Data Collection Agreement Form from Headway—that agreement is binding for the Mental Health Study as well. If you have any questions about this, discuss them with your FS.

2. EQUIPMENT AND MATERIALS

2.1 Introduction

For the screening and initial NSDUH interview, you will utilize your NSDUH equipment and materials. However, for respondents that have been selected for a follow-up interview, there are several additional materials you will use. You must use the correct version of the materials as appropriate during the MHSS. You will also need to be familiar with terminology specific to the Mental Health Study.

2.2 Equipment

The equipment for the MHSS will be the same equipment used for your NSDUH assignment. No changes have been made to the iPAQ screening program for the Mental Health Study. The NSDUH interview has been programmed so that the recruitment scripts for the follow-up interview will appear in the back-end CAPI portion when necessary.

2.2.1 Technical Support Procedures

Technical support for the MHSS will be provided by the Technical Support Group (TSG). For any technical issues, call your FS first, then, if necessary, call . Refer to ***Section 8.5*** in your ***FI Computer Manual*** for guidelines on contacting Technical Support.

2.3 Terminology

Terminology specific to the Mental Health Study is defined in this section. These terms will be used throughout the handbook, as well as during training and data collection activities. It is important that all members of the project (e.g. FIs, FSs, Technical Support personnel, etc.) understand and utilize these terms when discussing Mental Health Study data collection activities.

Mental Health Surveillance Study (or Mental Health Study, or MHSS): Nationwide study first conducted in 2008 and continuing this year to assess the NSDUH's Mental Health module questions. This study collects data from all adult interview respondents through a Mental Health module within the initial NSDUH interview, and through in-depth follow-up telephone interviews with selected adult respondents. For additional information refer to ***Chapter 1***.

Initial Interview: The NSDUH interview conducted with respondents, including respondents selected for the Mental Health Study follow-up interview. After completion of the initial interview, this will be your last contact with the respondent. Refer to ***Chapter 3*** for details.

Follow-up Interview: The follow-up telephone interview conducted with selected respondents who complete the initial interview. A telephone interviewer will complete the follow-up interview, which will ask questions about various mental health and other related issues.

Recruitment Scripts: After completing the initial interview, for selected respondents you will read a script programmed in the back-end portion of the interview that describes the follow-up interview to the respondent. These series of screens are referred to as "recruitment scripts" and it is the first time the follow-up interview is introduced to the respondent. ***Section 3.4*** contains the recruitment scripts.

2.4 Materials

In addition to the NSDUH materials used during the initial interview, for respondents selected for the follow-up interview there are several special Mental Health Study materials. These materials are described in this chapter as well as their location and color scheme. The use of each material will be discussed throughout this handbook.

The Mental Health Study materials include a statement about the respondent having been "randomly chosen for this special study" without providing sample size details. Zero, 1 or 2 residents from a household may be selected to complete the follow-up interview. Additionally, the materials indicate the follow-up interview questions focus on various mental health issues such as depression, anxiety, post traumatic stress disorder, and substance dependence rather than tobacco, alcohol, and drug use or non-use.

It is important to be organized and make sure you have everything you need when you go to the field to work, including enough money to cover NSDUH and Mental Health Study incentive payments. To assist you with materials organization, you will be given individual packets of Mental Health Study materials with your bulk supplies. Each packet will contain one copy of each Mental Health Study material in a large envelope for use during the recruitment scripts. These materials are the Follow-up Study Description, Follow-up Interview Payment Receipt, and Reminder Card. Keep two of these Mental Health Study materials packets in your laptop bag. After you use a packet of materials with a respondent selected for the follow-up interview, ask your FS to order a replacement packet. You should have two packets of materials at all times.

It is important that information about the follow-up interview be introduced when it is intended, during the back-end CAPI portion of the initial interview. **Under no circumstances should you mention this possibility to the respondent before this time**. This is why using the correct materials is so important. For example, if the Follow-up Interview Payment Receipt was used for the initial NSDUH interview, the information within the material would not be accurate for the initial interview.

Mental Health Study materials have not been translated into Spanish, because there will be no Spanish follow-up interviews completed on the Mental Health Study. All selected respondents must be able to complete the initial interview in English. If the initial interview is completed in Spanish, the respondent will not be eligible for the follow-up interview and the recruitment scripts will not appear.

At the conclusion of data collection, as directed by your FS, make arrangements to properly dispose of any remaining Mental Health Study materials by shredding and then either recycling or trashing the shredded documents.

2.4.1 Screening

There are no special Mental Health Study procedures or materials for the screening because it is not until the end of the initial interview that a respondent is selected and recruited to take part in the follow-up interview. Therefore, no changes have been made to the screening program for the Mental Health Study.

2.4.2 Initial Interview

The initial interview utilizes the 2012 NSDUH materials. However, for a respondent selected for the follow-up interview, there are several additional items you will use when reading the recruitment scripts. Listed below are the Mental Health Study materials you will use exclusively for the recruitment process. The table gives additional information about each material such as the color, the modification made to the form from the main study version, and the handbook exhibit number.

Mental Health Study Material	Color	Location	Modification	Exhibit #
Follow-up Study Description	Blue	Mental Health Study Materials Packet	• Title refers to follow-up interview • Description of study • Used during recruit process for follow-up interview at the end of the initial interview	**Exhibit 2.1**
Follow-up Interview Payment Receipt	Blue, 3 ply	Mental Health Study Materials Packet	• Title refers to follow-up interview • Given to the respondent along with the $30 payment after they agree to participate in the follow–up interview	**Exhibit 2.2**
Reminder Card	Blue	Mental Health Study Materials Packet	• Includes contact days & times as a reminder for the respondent	**Exhibit 2.3**

Exhibit 2.1 Follow-up Study Description

U.S. Public Health Service
Center for Behavioral Health
Statistics and Quality

Follow-up Study Description

You have been randomly chosen for this special study for the 2012 National Survey on Drug Use and Health. This study, sponsored by the United States Public Health Service, will ask questions about various mental health issues such as depression, anxiety, post traumatic stress disorder and substance dependence. Although there is no benefit to you personally, knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.

If you agree to participate in this follow-up interview, your first name and telephone number will be collected but will be used only for re-contact purposes. Your name and telephone number will not be included on the interview forms on which your answers will be written, or on any interview audio files that might be recorded. While the interview has some personal questions, federal law protects the privacy of your answers and requires us to keep all of your answers confidential. Any data that you provide will only be used by authorized personnel for statistical purposes according to the Confidential Information Protection and Statistical Efficiency Act of 2002. The only exception to this promise of confidentiality is if you tell the interviewer that you intend to seriously harm yourself or someone else; in this situation RTI may need to notify a mental health professional or other authorities.

The interview will be conducted over the phone and takes on average an hour to complete. Your participation is voluntary. You may consider some of the questions to be sensitive in nature and some of the questions also may make you feel certain emotions, such as sadness. Remember that you can refuse to answer any questions that you do not want to answer, and you can stop the interview at any time. If you become upset at any time during the interview and wish to speak to a mental health professional about how you are feeling, the interviewer can again provide you with the toll-free hotline numbers that are printed on your payment receipt from the first interview. **If you agree to complete the interview, you will receive $30 today.**

If you have questions about the study, call the Project Representative at . If you have questions about your rights as a study participant, call RTI's Office of Research Protection at (a toll-free number). You can also visit our project Website: http://nsduhweb.rti.org/ for more information.

Thank you for your cooperation and time.

, Project Officer
Center for Behavioral Health Statistics and Quality
Substance Abuse and Mental Health Services Administration (SAMHSA)
U.S. Public Health Service
Department of Health and Human Services

Exhibit 2.2 Follow-up Interview Payment Receipt

Follow-up Interview Payment Receipt

United States Public Health Service

and

Research Triangle Institute

thank you for agreeing to participate in a special study for the 2012 National Survey on Drug Use and Health.

In appreciation of your participation in this important study, you are eligible to receive a $30 cash payment.

Since maintaining the confidentiality of your information is important to us, your name will not be entered on this form. However, the interviewer must sign and date this form to certify you received (or declined) the cash payment.

_____	_____	__ __ __ __ __ __ __ __ __ __ - __
Interviewer	Date	Case ID

☐ Accepted Cash Payment ☐ Declined Cash Payment

If you ever feel that you need to talk to someone about mental health issues, you can call the National Lifeline Network. Counselors are available to talk at any time of the day or night and they can give you information about services in your area.

1-800-273-TALK or 1-800-273-8255
1-888-628-9454 (Spanish)
http://suicidepreventionlifeline.org/

If you ever feel that you need to talk to someone about drug use issues, you can call the Substance Abuse and Mental Health Services Administration's Treatment Referral Helpline. This is a 24-hour service that will help you locate treatment options near you.

1-800-662-HELP or 1-800-662-4357
1-800-487-4889 (TDD)
http://findtreatment.samhsa.gov

Disposition: Top copy to Respondent, yellow to Field Supervisor, pink to Field Interviewer.

Exhibit 2.3 Reminder Card

We appreciate you taking time for this important study and look forward to speaking with you soon.

Your suggested contact days and times are:

Day	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Time							

Research Triangle Institute
Research Triangle Park, NC 27709-2194

3. INITIAL INTERVIEW

3.1 Introduction

There are many necessary components that contribute to the overall success of every scientific study. You are responsible for one of the most important aspects of the Mental Health Study—making sure the initial NSDUH interview is administered properly according to established survey procedures. This chapter discusses the steps required to complete the initial interview and administer the recruitment scripts for the follow-up interview when they appear on the laptop screen.

3.2 Initial Interview

For research purposes, it is essential to treat all interview respondents in the same manner. The initial interview is the same for all NSDUH respondents. Adult respondents completing the initial interview in English may be randomly selected for participation in the Mental Health Study. For selected respondents, they will be asked to complete an additional follow-up telephone interview (also in English) after the initial interview is completed (see ***Chapter 1***). Since you will not know which respondents will be selected for the follow-up interview until the recruitment scripts appear, it is crucial that you have and use the correct forms during the initial interview. Reference ***Section 2.4.2*** for a list of exclusive Mental Health Study materials used to complete the follow-up interview recruitment process.

3.3 Incentive Procedures

Respondents receive a $30 cash payment for completing the initial interview. The procedures for providing the initial incentive payment are the same, including the use of the Interview Payment Receipt. However, selected interview respondents who agree to complete the follow-up telephone interview will be given an additional $30 when prompted by the recruitment scripts at the end of the initial interview. Therefore, there is the possibility of paying a selected respondent a total of $60 during your interview encounter. Always be sure you have enough incentive money with you in case a respondent is selected and agrees to complete the follow-up interview. Since a household may have two selected respondents, you must have enough incentive money on hand ($120) to pay incentives to both respondents. As needed, replenish your supply by visiting a bank or ATM before contacting another household. See ***Section 3.4*** for more information.

3.4 Recruitment Process

The recruitment scripts are the first opportunity for you to tell the respondent about the possibility of participating in a follow-up interview and receiving an additional $30 for their agreement to participate. For information on how to answer respondent questions about the follow-up interview, see ***Chapter 4***.

The beginning of the recruitment process, where you introduce the follow-up interview to the respondent, is done in a standardized way through a series of back-end CAPI screens. <u>You must read these recruitment scripts verbatim</u>.

The recruitment process begins after you pay the respondent for the initial interview (INCENT01). The next screen (RECRUIT1) introduces the study to the respondent and gives a brief explanation of the process, including the time frame for the follow-up interview and the additional cash payment. You will also give the respondent a copy of the Follow-up Study Description which provides more information on the telephone interview. It is imperative to give the correct version of the study description at this point to make sure the respondent receives all necessary information for making an informed decision about participation.

For respondents who agree to participate, enter "1" on this screen. The next screen to appear is shown on page 3-6.

If the respondent agrees to participate but is not available during the specified time period (within the next two weeks), enter "3" on the RECRUIT1 screen and the next screen appears—RECRT4WK. This screen informs the respondent an interviewer will be available to complete the interview within the next four weeks to accommodate their schedule.

Telephone interviewers have up to four weeks to complete the follow-up interview. However, our preference is to complete the follow-up interview with the respondent in the first two weeks. This screen allows respondents who are not available during the first two weeks to participate if they will be available during the four week time period. Never suggest the option of completing the follow-up interview within four weeks rather than two weeks until seeing this screen.

If the respondent refuses to participate in the follow-up interview, do your best to address any questions the respondent has at that time to try to persuade him/her to participate. However, if the respondent continues to refuse, you are not to make a return visit to the DU to try to convert the follow-up interview refusal. An initial refusal to the follow-up interview is considered a final refusal. In this situation, enter "2" on either the RECRUIT1 or RECRT4WK screens as appropriate, and the next screen, REFFEAS, will appear asking the respondent why they do not want to participate. This has an open-ended answer field which allows you to enter up to 100 characters of text. Enter the answer clearly and concisely.

The next screen, THANKR2, appears where you should politely thank the respondent for his/her time and go on to finish the FI debriefing questions and wrap-up the interview just as you normally do.

Forms Answer Navigate Options

Thank you for your time. I just need to finish a few questions on my own to show that I did the interview. This will only take a few minutes.

BE SURE YOU HAVE YOUR SHOWCARD BOOKLET, CALENDAR, QC ENVELOPE W/ FORM, AND PAYMENT RECEIPT COPIES.

PRESS [ENTER] TO CONTINUE.

THANKR2

If the respondent is available and agrees to complete the follow-up interview, enter "1" on the RECRUIT1 or RECRT4WK screens and you will be routed to the RECRUIT2 screen shown below. Here you obtain the respondent's first name and telephone number so the telephone interviewer will be able to contact the respondent for the follow-up interview.

To enter the respondent's contact information, simply type the appropriate information into the field, starting with the first name, and press [**ENTER**] after each entry to move to the next field. If you need to back up to a previous field, press the up and down arrows on the keyboard to move to the correct field. Additional information about the first name or telephone number provided by the respondent should be entered in the Notes field.

It is extremely important to get the correct phone number for the respondent as this encounter will be the last opportunity for face-to-face contact with this person. Therefore, to ensure you collect accurate contact information, ask the respondent to repeat the telephone number given. You are also instructed to read the contact information you have entered into the computer back to the respondent for confirmation as needed.

Follow the script to ask the respondent for an additional contact number, entering this information in the Notes. Include a comment describing the type of number, such as cell phone or work phone.

Next, you will obtain the best days and times to contact the respondent within the next two or four weeks (dependent upon the respondent's recorded availability) on the RECRUIT3 screen. Previously, the RECRUIT1 or RECRT4WK screen indicated that the first telephone call would be to schedule a convenient time to complete the full interview, but the respondent can complete the interview at that time as well. This statement should clarify that the times they are providing for re-contact are not necessarily firm interview appointment times but rather times for the telephone interviewer to contact them to schedule a mutually convenient time for the follow-up interview.

You may need to probe for several good days and times. For example, if a respondent says the best time to reach him is on Sundays after 5 pm, enter this information and then ask what other days and times would be convenient. Telephone interviewers are available on evenings and weekends as well as during weekdays to accommodate the respondent's schedule. As needed, share this information with the respondent to obtain additional days and times to contact them.

Once you enter the best days and times for re-contact, read the information you typed into the laptop back to the respondent to confirm you have entered it correctly. Then, you are instructed to complete a Reminder Card and hand it to the respondent. Under the listed days on the card, write the times the respondent provided for contact. Give the card to the respondent to use as a reminder of the suggested days and times they may be contacted about the follow-up interview, then read the text informing them that the interviewer may try to reach them at other times as well.

An example of the RECRUIT3 screen referencing the four week time period is shown below. This screen will appear for respondents who indicate they are available during the four week time period on RECRT4WK.

Forms Answer Navigate Options

Please also let me know the best days and times when you will be available in the next four weeks. I will give this information to the interviewer, and he or she will try to contact you during one of these times.

ENTER BEST DAYS/TIMES. AS NEEDED, PROBE FOR ADDITIONAL BEST DAYS/TIMES.
READ THE INFORMATION ENTERED TO THE RESPONDENT AND CONFIRM IT IS CORRECT.

COMPLETE A REMINDER CARD AND HAND TO THE RESPONDENT.

I have entered these days and times in the computer and recorded them on this card. Please note the interviewer may try to reach you at other times as well.

INTERVIEWER NOTE: ADDITIONAL INFORMATION REGARDING THE BEST DAYS OR TIMES PROVIDED BY THE RESPONDENT SHOULD BE ENTERED IN THE NOTES FIELD. YOU MAY ENTER UP TO 50 CHARACTERS.

TELEPHONE INTERVIEWERS ARE AVAILABLE EVENINGS AND WEEKENDS.

PRESS [ENTER] TO CONTINUE.

RECRUIT3 - BEST DAYS	
BEST TIMES	
NOTES	

Next, the INCENTMH screen will appear. At this point you should pay the respondent the additional $30 for the follow-up telephone interview and sign the Follow-up Interview Payment Receipt, reading the scripted text and following the instructions on the screen. This screen is similar to the earlier incentive screen that directs you to pay the respondent for completion of the initial interview. As with the initial interview payment, it is imperative to follow the instructions in the order listed and to read the text verbatim. As stated earlier, you must have enough incentive money on hand ($120) in case two respondents are selected for the follow-up interview.

After paying the respondent the $30 for the follow-up interview, end the recruitment process by reading the THANKR2 screen. At this point, you should thank the respondent for his/her time and then finish any the remaining interview tasks, including the FI debriefing questions.

If you would like to review or practice completing the recruitment process, a list of Mental Health Study initial interview practice cases can be found in ***Exhibit 3.1***.

3.5 Transmit Each Day You Work

After completing the initial interview and returning home, you must transmit to RTI. Telephone interviewers work on a very tight schedule in order to complete the follow-up interview within the four-week deadline and preferably within the first two weeks. Once the follow-up interview information is received at RTI, the case is assigned to a telephone interviewer within 24 hours. Please note that depending on when the case is transmitted, the telephone interviewer may not be able to call the respondent the very next day.

Exhibit 3.1 Mental Health Study Initial Interview Practice Cases

The following cases may be used for Mental Health Study practice. Each case has one practice interview denoted (Interview A). This chart includes all the information you need to complete the interview.

Please note: These cases will advance from the CONFIRM screen in the front-end Demographics to Question Q124 which asks about the number of telephones in the respondent's household. This will allow you to easily practice the recruitment portion of the interview.

CASE ID	QUEST ID	QUALITY CONTROL ID
ZZ10010025A	9995482	X10-0008
ZZ10010027A	9996016	X10-0016
ZZ10010029A	9995491	X10-0024
ZZ10010031A	9996025	X10-0032
ZZ10010033A	9995507	X10-0040
ZZ10010034A	9996034	X10-0057
ZZ10010035A	9995516	X10-0065
ZZ10010036A	9996043	X10-0073
ZZ10010037A	9995525	X10-0081
ZZ10010039A	9996052	X10-0099

This page intentionally left blank.

4. OBTAINING PARTICIPATION

4.1 Answering Respondent Questions

When answering a respondent's questions, at any phase—screening or initial interview, it is important that nothing you say affect the respondent's answers.

One of the most important protocols of the Mental Health Study is that respondents not know about the possibility of a follow-up telephone interview until the recruitment screens appear at the end of the initial interview. **NEVER** mention the follow-up interview until you see the recruitment scripts. **NEVER** use the possibility of an additional interview payment of $30, while trying to persuade the respondent to complete the screening or initial interview.

What can you say to the respondent about the follow-up interview? If the respondent asks you why he/she has been selected to complete an additional interview, it's important to convey the purpose provided in the recruitment scripts and Follow-up Study Description:

> **"You have been randomly chosen for this special study for the National Survey on Drug Use and Health. This study, sponsored by the U.S. Public Health Service, will ask questions about various mental health issues. Knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States."**

It would be useful to memorize this statement in case a respondent asks you about this. This scripted response is also provided on the RECRUIT1 screen (***Exhibit 4.1***). However, only offer this information if the respondent questions you. It is important to respond using these words as opposed to something else that seems neutral, but may be biasing. We do not want the respondent to feel they were selected for the follow-up interview because of the answers they provided during the initial interview. Therefore, it is important to emphasize the respondent was randomly chosen for the follow-up interview.

Like the text of the Follow-Up Study Description, the above response refers to a 'special study' instead of the Mental Health Study, as the name may raise concerns for some respondents. Use the terms 'special study,' 'telephone interview,' 'follow-up study,' or 'important study' with respondents.

If a respondent asks why a different interviewer has to contact them by phone for the next interview, tell the respondent:

> **"A different interviewer is required to do the second interview because they have been specifically trained on the follow-up interviewing process."**

If a respondent expresses concerns that they only have a cell phone as a method of contact or to complete the follow-up interview, explain the telephone interviewer will call them to set up a convenient time to complete the interview, such as during the evening or on the weekend when cell phone minutes are often free. The telephone interviewer can also arrange to call them at an alternate phone number for the follow-up interview, such as a work number, rather than their cell phone.

For more information about answering respondent questions, including questions related to telephone use, see ***Exhibit 4.4***.

4.2 Refusals

If a respondent refuses to participate in either the initial interview or follow-up interview, address the respondent's questions and concerns using your knowledge and the appropriate study materials.

4.2.1 Initial Interview Refusals

For initial interview refusals, do your best to convert the respondent using the appropriate study materials (see ***Chapter 2***). Once again, **NEVER** mention the possibility that if the respondent completes this interview, he/she can also do a follow-up telephone interview for an additional $30 incentive. Although every FI should strive for good response rates, it is more important that the correct Mental Health Study protocols are followed.

4.2.2 Follow-up Interview Refusals

If the respondent completes the initial interview, but refuses to agree to complete the follow-up interview, address any questions the respondent has at that time, but do not return to the DU to try to convert the follow-up interview refusal. An initial refusal to the follow-up interview is considered a final refusal.

During the recruitment process, the objection by the respondent may take place while on the RECRUIT1 (***Exhibit 4.1***) or RECRT4WK (***Exhibit 4.2***) screens. If you are unable to convert the respondent's refusal, enter "2—RESPONDENT DOES NOT AGREE TO RECONTACT."

A respondent may also decide against participating while on a later recruitment screen, perhaps while providing a phone number or best days or times. In this situation, press F9 as needed to return to RECRUIT1, and enter "2—RESPONDENT DOES NOT AGREE TO RECONTACT."

You will then be routed to the REFFEAS (***Exhibit 4.3***) screen which prompts you to ask the respondent why he/she does not want to participate. This question has an open-ended answer field and will allow up to 100 characters of text. Be as clear and concise as possible.

Remember, even if the respondent refuses the follow-up interview, you must enter a code 70 into the iPAQ after completing the initial interview. You do not need to enter additional ROCs for the follow-up interview as that information is gathered in the recruitment screens. Your work on the case is considered complete after finishing the initial interview and attempting to recruit the respondent for the follow-up interview.

4.3 Other Non-response

Be familiar with how to handle these types of non-response.

4.3.1 Respondent Unavailable

A respondent may not be able to complete the follow-up interview when you read the RECRUIT1 (***Exhibit 4.1***) or RECRT4WK (***Exhibit 4.2***) screens because they are unavailable during the specified time frame—that is, within two or four weeks, respectively. In this situation, enter "3—RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD." Again, even if the respondent is unavailable to complete the follow-up interview, you must enter the code 70 into the iPAQ after completing the initial interview.

4.3.2 No Phone Access

If a respondent would like to participate but does not have access to a phone, probe to see if there is another phone number where they could be reached. As needed, suggest a cell or work phone, but do NOT mention specific locations such as a neighbor, friend or relative. If the respondent is not able to provide a phone number to contact them, explain that in order to participate in the study they must be able to complete the follow-up interview by telephone. At the RECRUIT1 screen, enter "2—RESPONDENT DOES NOT AGREE TO RECONTACT." Next, at the REFFEAS screen, enter a brief response to document the respondent does not have access to a phone. Only for situations where the respondent does not have access to a phone, you are not required to read the REFFEAS screen to the respondent.

Exhibit 4.1 RECRUIT1 Screen

Forms Answer Navigate Options

You have been randomly selected to participate in one additional study for the U. S. Public Health Service. This interview will ask questions about mental health issues. It will be conducted over the telephone and will take about an hour. Participation in this interview is voluntary and all of your answers will be kept confidential.

HAND FOLLOW-UP STUDY DESCRIPTION TO RESPONDENT. Please read this statement. It describes the survey and the legislation that assures the confidentiality of any information you provide.

If you agree to participate, I will pay you an additional $30 today. Within the next two weeks, a different interviewer will call you to explain more about the interview and to schedule a convenient time to complete it. If you wish, you may complete the full interview when the interviewer calls.

IF ASKED "WHY WAS I SELECTED?" (You have been randomly chosen for this special study for the National Survey on Drug Use and Health. This study, sponsored by the U.S. Public Health Service, will ask questions about various mental health issues. Knowledge gained from this study will improve our ability to describe and understand mental health issues in the United States.)

1. RESPONDENT AGREES TO RECONTACT
2. RESPONDENT DOES NOT AGREE TO RECONTACT
3. RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD

RECRUIT1

Exhibit 4.2 RECRT4WK Screen

Forms Answer Navigate Options

To accommodate your schedule, an interviewer will be available to call you about this study and schedule a convenient time to complete the interview within the next four weeks.

- 1. RESPONDENT AGREES TO RECONTACT
- 2. RESPONDENT DOES NOT AGREE TO RECONTACT
- 3. RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD

RECRT4WK

Exhibit 4.3 REFFEAS Screen

Exhibit 4.4 Respondent Questions and Responses

Respondent Question	FI Response
"What types of questions will the interviewer ask me?"	"The study asks questions about various mental health issues such as depression, anxiety, post traumatic stress disorder and substance dependence."
"How long will the telephone interview last?"	"The interview takes about an hour."
"Why was I selected for an additional study?"	"You have been randomly chosen for this special study for the National Survey on Drug Use and Health." OR "You were randomly chosen to participate in this special study. The computer program makes the selection and I have no idea who will be chosen until it appears on the screen."
"Why must there be a different interviewer? I'd like to do the interview now."	"A different interviewer is required to do the second interview because they have been specifically trained on the follow-up interviewing process."
"How will the interviewer's phone number display on Caller ID?"	"Each interviewer's phone number will be different. It will not show up as an RTI number since the interviewers work in different locations across the country."
"Why is my participation so important?"	"You are important because knowledge gained from the study will improve our ability to describe and understand mental health issues in the U.S."
"Why am I getting $60 when my brother only got $30 for the interview?" (for 2 person households where only one person is selected for the Mental Health Study)	"You were randomly chosen and agreed to participate in this special study for the National Survey on Drug Use and Health. Since the selection process is random, only a limited number of people will be asked to complete the additional interview."
"Can I complete the interview in Spanish?" (or any other language) (for bilingual respondents who complete the initial interview in English, but would like to complete the follow-up interview in Spanish or another language)	"For this study, you must be able to complete the entire interview in English."

Exhibit 4.4 Respondent Questions and Responses (continued)

Respondent Question	FI Response
"Why do you need a second phone number?"	"Providing a second phone number assists the interviewer in contacting you about the follow-up interview. While we prefer to have a second phone number available, you are not required to give us another number."
"I don't have a phone. Can I still do the interview?"	"To participate in this study, you must be able to complete the interview by telephone. Is there another phone number where we could reach you?" (See ***Section 4.3.2*** for more information)
"I only have a cell phone with very limited minutes."	"The telephone interviewer will call you to set up a convenient time to complete the interview, such as during the evening or on the weekend when cell phone minutes are often free." OR "The telephone interviewer can call you at an alternate number for the interview, such as a work number."
"Can I do the interview on my cell phone as I drive to work?"	"The telephone interviewer would not want you to be distracted while driving, but can arrange for another convenient time to complete the interview that works for your schedule."
"How many people are chosen to complete the (telephone) interview?"	"Our goal is to complete 1,500 interviews this year."
"I don't want to do an interview over the phone for hour. Can someone interview me in person?"	"A different interviewer is required because they have been specifically trained on the follow-up interviewing process. There are a limited number of these interviewers available, and they are only able to complete the interview by telephone."

5. ADMINISTRATIVE PROCEDURES

5.1 Introduction

Your documentation and reporting of field activities allows you, your FS, and the RTI project management staff to closely monitor NSDUH and Mental Health Study data collection efforts. Changes, additional procedures, and important clarifications are noted below.

5.2 Project Charge Code

All Mental Health Study field work should be charged to the 2012 NSDUH project number, which is 0212800–001.106.002.

5.3 Completing and Submitting Weekly ePTE

You should maintain an ePTE working copy that combines NSDUH and Mental Health Study hours, miles, and expenses. There is no need for you to keep separate records of charges. In the Notes field of the ePTE where you list the Case IDs of completed initial interviews, add an "MH" in parentheses to the end of Case IDs that generated a follow-up interview. You must transmit a complete and accurate ePTE report each week by 10:30 pm (EST) on Sunday from your laptop. For general ePTE questions, refer to ***Chapter 11*** of your ***FI Manual***.

5.4 PT&E Summary Data

Enter your weekly PT&E Summary Data in your iPAQ so that your entry includes NSDUH and Mental Health Study work. You should not enter or transmit weekly PT&E Summary Data separately for the Mental Health Study.

5.5 Incentive Advances

You are responsible for paying respondents cash incentives and monitoring your incentive funds on a weekly basis. You will not be issued a separate incentive advance for the Mental Health Study. Instead, pay respondents who agree to the follow-up interview using your 2012 NSDUH Incentive Advance.

You are required to pay all selected interview respondents who agree to the follow-up interview in cash. Therefore, when cashing the Headway incentive check, request that it be dispensed in ten and twenty dollar bills. A respondent who is selected and agrees to be contacted for a follow-up telephone interview has the opportunity to receive a total of $60 at the completion of the initial interview via two separate incentive payments.

A respondent who completes the initial interview receives $30 for the completion of that initial interview. If they then agree to be re-contacted for the follow-up telephone interview, they receive an additional $30 cash incentive. Refer to ***Chapter 3*** of this handbook for more details. Headway reimburses all incentives paid based on the total amount of incentives entered in Column 16 of your ePTE. As a result, your incentive cash amount will always equal the original amount. A summary of incentives paid must be included in the Notes section of the ePTE, as well as an explanation for any case in which the respondent refused to accept the cash incentive, including the CaseID. For example, the correct way to

summarize incentives paid for 6 completed NSDUH interviews, one of which yielded a follow-up interview would be: "6 int. @ $30, 1 MH @$30." See ***Exhibit 5.1*** for an example of a properly completed ePTE working copy, including a follow-up interview.

1.1.1 5.5.1 Incentive Advance Balance Sheet

Keeping track of incentive advances received, incentive payments made to respondents and reimbursements from Headway requires careful and detailed records. Use the Incentive Advance Balance Sheet to help you keep a record of incentive payments and Headway reimbursements (see ***Exhibit 11.6*** in the ***FI Manual***). Your Headway Advance Balance Booklet contains instructions on how to monitor your incentive advance money and also a supply of blank Incentive Advance Balance Sheets. Additionally, instructions for repaying incentive advance balances are included in ***Section 11.7.3*** in the ***FI Manual***.

5.6 Transmitting to RTI

Each day you work, be sure to transmit to RTI. It is essential that all NSDUH work is received on a timely basis, but even more so for the Mental Health Study because of the deadline for completion of the follow-up interviews. Telephone interviewers must complete the follow-up interview within four weeks, and preferably within the first two weeks. This tight schedule means you must transmit each night you work to ensure information about the follow-up interview is received promptly at RTI.

Exhibit 5.1 Properly Completed ePTE Working Copy

WORKING COPY of the ePTE

Data Collection ePTE

Week Begin 1 / 08 / 2012

Travel (circle one) Yes / No

Project # 0212800-001.106.002

Supervisor Kristen Effess

Area # 999

Day of Week	Date	Production: # Comp. Interviews*	Time: Training	Home Study	Travel	Contact Locate	Inter-viewing	Editing	Confer-ence	Other^ (include Trans Time)	Daily Total@
Sun	1-08										
Mon	1-09	2			1.00	4.00	2.50			0.25	7.75
Tues	1-10										
Wed	1-11	1			1.00	3.00	1.50		0.75	0.25	6.50
Thur	1-12										
Fri	1-13										
Sat	1-14	3			1.25	1.00	4.50			0.25	7.00
TOTALS		6			3.25	8.00	8.50		0.75	0.75	21.25

Notes

XX10010035A
XX11140306B
XX10010048A (MH)
XX11140021A
XX1114A028A
XX10010116B

Other time on 1-09, 1-11, and 1-14 for transmitting

6 int @ $30
1 MH @ $30

$0.44 expense on 1-15 for postage

*Record Case ID# for each completed IVW.
^Include Transmitting time in "Other." Explain all "Other" entries in Notes.
Goes on Expense Report with receipt attached.
$Expenses over $10 go on Headway Expense Report.
!Include summary of incentives paid (ex: # int. @ $30, # MH @ $30)
@ ePTE program will total for you.

Day of Week	Date	Expense: Miles Driven	Lodging#	Meal Expenses	Tele-phone#	Auto Rental#	Miscellaneous$	Incentive!	Total Expense@
Sun	1-08								
Mon	1-09	50						60	0
Tues	1-10								
Wed	1-11	60						60	0
Thur	1-12								
Fri	1-13								
Sat	1-14	75					0.44	90	$0. 44
TOTALS		185					0.44	210	$0.44

NOTE: Remember to note your Total Hours, Total Miles Driven, & Total Expenses for entry into the iPAQ PT&E Summary

DAILY NOTES SPACE:

	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
Odometer START		40,100		40,210			40,350
Odometer STOP		40,150		40,270			40,425
Time START		11:00 am		2:00pm			10:00am
Time STOP		2:30 pm		7:30pm			2:00pm
Time START		3:00 pm					5:30pm
Time STOP		7:00 pm					8:15pm

This page intentionally left blank.

6. QUALITY CONTROL

6.1 Introduction

The importance of quality work at every point in the process of collecting research data cannot be emphasized enough. As interviewers, you are the direct link in the flow of information for the survey and you must carefully perform the screening and interview procedures as instructed so the results are of the highest possible quality.

6.2 Data Quality Reminders

The Mental Health Study protocols must be followed precisely to ensure the validity of this special study. If you focus on the data quality reminders listed below, you should be able to complete your work successfully and without data quality problems.

1. Use the correct field materials for the initial interview and recruitment scripts (refer to ***Chapter 2*** as needed).

2. Read all screens verbatim. Since you will be recruiting so few respondents for the follow-up interview, you may be initially surprised when the recruitment scripts appear. Simply trust the instrument and read the screens verbatim following all the instructions. The screens tell you exactly what to do each step of the way.

3. Never mention the possibility of a follow-up interview until scripted in the back-end CAPI of the initial interview.

6.3 Verification

The Mental Health Study procedures for verification are the same as the 2012 NSDUH.

6.4 Ongoing Training

This handbook should be carefully studied as you prepare for fieldwork each quarter in 2012, and referred to during data collection as questions arise. Keep it in your laptop bag so it is available, but be careful to not show it to respondents, either on purpose or accidentally.

You will complete a separate iLearning course before the start of each quarter in 2012 as a refresher and further training on Mental Health Study procedures. You must store your iLearning CD in a safe place at home. Do not store the CD in your laptop bag or car as damage could easily occur.

As always, should you have any questions about the study not covered in this handbook, be sure to ask your FS.

This page intentionally left blank.

Appendix F: New Clinical Interviewer Training Agenda

2011 MENTAL HEALTH SURVEILLANCE STUDY: NEW CLINICAL INTERVIEWER TRAINING AGENDA

DAY 1	8:15 – 5:00	**Objective:** Introduce the study, the equipment and systems used

8:15 – 9:15 **1. INTRODUCTION TO THE STUDY (1 hour)**

- Staff introductions, including ice breaker
- Review training schedule, training courtesies
- Introduction to the handbook
- Explanation of NSDUH, MHSS/EMHSS
  - VIDEO – MHSS Recruitment Process [5 minutes]
- Overview of CI responsibilities; certification process

9:15 – 10:00 **2. INTRODUCTION TO THE EQUIPMENT (1 hour, 15 minutes)**

- Issue equipment, sign EARFs
- Overview of the Laptop (hardware, software)
- Calling card
- Computer Set Up
- RTI Webmail (and practice)

10:00 – 10:15 BREAK

10:15 – 10:45 **2. INTRODUCTION TO THE EQUIPMENT, cont.**

- CMS introduction
- Audio Recording (Demo with recording equipment)
- Care of Laptop, Technical Support Procedures

10:45 – 11:00 **3. CASE ASSIGNMENT & CMS FUNCTIONS (15 minutes)**

- CMS functions
- Receiving case assignment information

11:00 – 12:00 **4. CONTACTING RESPONDENTS (1 hour, 45 minutes)**

- Respondents' rights and confidentiality
- Review materials
- Contacting the Respondent
- Introducing the Study

12:00 – 1:00 LUNCH

1:00 – 1:45 **4. CONTACTING RESPONDENTS, cont.**

- Interviewing the Respondent
- Nonresponse

1:45 – 2:30 **5. DOCUMENTING RESULTS (45 minutes)**

- Use and importance of Record of Calls (ROCs)
- Status Codes
- Documenting ROCs, including comments
- Practice entering codes for various scenarios into CMS
- Status Code Exercises

2:30 – 3:00 **8. GENERAL CLINICAL INTERVIEWING SKILLS (30 minutes)**

TASK: Trainer will review probing, open-ended questions, anchoring, rating, ranking, etc.

CI MATERIAL: Blank SCID (for notes)

CI HANDOUT: SCID Documentation Handout

3:00 – 3:15 BREAK

3:15 – 3:30 **6. DISTRIBUTION OF MATERIALS (15 minutes)**

- Distributing materials
- Transfer of Audio Recordings
- Shipment of SCID; preparing FedEx Shipments

3:30 – 5:00 **7. ADMINISTRATIVE PROCEDURES (1 hour, 30 minutes)**

- ePTE Procedures (enter default values, practice with ePTE system, prepare ePTE for trip to training, exercises for dividing charges)
- Reporting to Data Collection Manager
- Reporting to Clinical Supervisor
- End of Day, review homework assignment

6:00 – 8:00 **"VIRTUAL" CLINICAL INTERVIEWER LAB** (Trainers on call to assist with homework)

HOMEWORK: Practice recording and uploading audio files
Carefully review Chapter 5 and Appendix D of Handbook

DAY 2	8:15 – 5:00	**Objective:** Focus on interviewing skills and operational definitions of each symptom

8:15 – 9:00 **8. SCID TRAINING - OVERVIEW** (Lecture/Demonstration)

TASK: Trainer will review the SCID Overview followed by a demonstration with CS as respondent and the new CI's conducting the interview in a round-robin.

TRAINER MATERIAL: SCID #1 with responses for demonstration

CI MATERIAL: Blank SCID (for notes and recording responses)

NOTE: The above task and materials list applies to all Lecture/Demonstrations during Day 2.

9:00 – 9:05 **SCID Screening** (Lecture/Demonstration)

9:05 – 10:00 **Module A – Depression** (Lecture/Demonstration)

10:00 – 10:15 BREAK

10:15 – 10:40 **Module A (cont.) – Mania and Mood D/O Due to GMC/Substance** (Lecture/Demonstration)

10:40 – 11:05 **Module A (cont.) – Dysthymia** (Lecture/Demonstration)

11:05 – 11:35 **Module B/C – Psychotic Screen, Module D – Mood Differential** (Lecture/Demonstration)

11:35 – 12:00 **Module E – Anxiety Disorders: PTSD** (Lecture/Demonstration)

12:00 – 1:00 LUNCH

1:00 – 2:15 **Module E – Anxiety Disorders cont.: Other Anxiety Disorders** (Lecture/Demonstration)

2:15 – 2:40 **Module F – Eating Disorders** (Lecture/Demonstration)

2:40 – 2:50 **Module G– Intermittent Explosive Disorder** (Lecture/Demonstration)

2:50 – 3:05 BREAK

3:05 – 3:45 **Module H – Substance Use Disorders** (Lecture/Demonstration)

3:45 – 3:55 **Module J – Adjustment Disorder** (Lecture/Demonstration)

3:55 – 4:15 **Module K – End of Interview (GAF)** (Lecture)

4:15 – 4:45 **Exercise**: GAF Scoring (Practice)

TASK: Individual Practice GAF scoring for 6 cases

TRAINER HANDOUT: GAF Scoring Exercise

CI HANDOUT: GAF Scoring Exercise

4:45 – 5:00 **End of Day** – Wrap-up

TASK: Answer questions. Hand out homework

CI MATERIAL: Homework handouts (GAF)

6:00 – 7:00 **CLINICAL INTERVIEWER LAB** (Optional)

HOMEWORK: **HOMEWORK: GAF SCORING**

TASK: Read 24 vignettes and practice GAF scoring

TRAINER HANDOUT: GAF Practice Exercise for Trainers (with GAF Answers in the back)

CI HANDOUT: GAF Practice Exercises

DAY 3	8:15 – 5:00	**Objective:** Review the structure of the SCID, skip patterns and putting it all together

8:15 – 8:30 **Review of Agenda and Day 2**

TASK: Collect GAF homework from CIs. Other trainers will grade responses and return to CIs later in the day

8:30 – 8:50 **Review of SCID Structure** (Lecture)

CI MATERIAL: Blank SCID (from Day 1 for notes)

8:50 – 9:50 **10. HANDLING UNUSUAL SITUATIONS (1 hour)**

- Cognitive impairment protocol and reporting
- Distressed Respondents

9:50 – 10:05 BREAK

10:05 – 10:15 **Calibration & Inter-rater Reliability** (Lecture)

10:15 – 11:45 **Exercise: Independent Rating of Recorded SCID**

INDEPENDENT RATING/GROUP CALIBRATION EXERCISE

TASK: Group will independently score a 90-minute real SCID interview (Q4 2010 IR Calibration Interview 4013551), followed by a 45-minute discussion of how the symptoms were coded.

TRAINER/CI MATERIAL: Blank SCID

11:45 – 12:45 LUNCH

12:45 – 1:30 **Exercise: Group Calibration** (Discussing SCID Ratings)

TASK: Group will discuss how the symptoms were coded.

TRAINER MATERIAL: Answer Sheet for Calibration Exercise

CI MATERIAL: Independently coded SCID

1:30 – 1:45 **Module K – End of the Interview (continued) and Module X – Interviewer Debriefing Section** (Lecture)

1:45 – 2:00 **Threats to Validity** (Lecture)

2:00 – 3:00 **Full Interview Exercise #1**

TASK: CIs will work in pairs to practice administering entire SCID. Trainers will float and provide supervision.

TRAINER/RESPONDENT MATERIAL: SCID #2 script for Full Interview 1

CI MATERIAL: Blank SCID

3:00 – 3:15 BREAK

3:15 – 4:15 **Full Interview Exercise #2**

TASK: CIs will work in pairs to practice administering the entire SCID. Trainers will float and provide supervision.

TRAINER/RESPONDENT MATERIAL: SCID #3 script for Full Interview 2

CI MATERIAL: Blank SCID

4:15 – 4:45 **Review** – Practice Interviews #1and #2 (Lecture)

4:45 – 5:00 **End of Day** – Wrap-up

CI HANDOUT: Script/Answer Sheet for Full Interview Exercises #1 and #2

CI HANDOUT: Graded GAF Practice Exercises and Answer Sheet

6:00 – 7:00 **CLINICAL INTERVIEWER LAB** (Optional)

HOMEWORK: Review materials from Days 2 and 3
Prepare any written questions for Day 4

DAY 4	8:15 – 4:15	**Objective:** Practice the entire interview process – from case assignment to shipment

8:15 – 8:30 **REVIEW OF DAY 3**

- Answer questions

8:30 – 9:30 **Review of GAF Scoring Homework**

HOMEWORK #1: GAF SCORING

TASK: Review GAF scores to the 24 vignettes

TRAINER HANDOUT: GAF Practice Exercise for Trainers (with GAF Answers in the back)

9:30 – 10:00 **9. SCID PAIRED MOCK (4 hours, 45 minutes)**

CIs go through the entire process of getting the case, making an appointment, conducting the interview, packaging SCID for shipment, etc.

Trainers will float and provide supervision.

RESPONDENT MATERIAL: SCIDs #4 and #5 for Paired Mocks 1 and 2

CI MATERIAL: Blank SCID

10:00 – 10:15 BREAK

10:15 – 12:00 **9. SCID PAIRED MOCK, cont.**

- Paired Mock exercise (continued)

12:00 – 1:00 LUNCH

1:00 – 2:30 **9. SCID PAIRED MOCK, cont.**

- Paired Mock exercise (continued)

2:30 – 2:45 BREAK

2:45 – 3:45 **9. SCID PAIRED MOCK, cont.**

- Review of Mock SCIDs (Lecture)

CI HANDOUT: Script/Answer Sheet for Paired Mock Exercises

3:45 – 4:15 **11. WRAP-UP (30 minutes)**

- Summary/Answer CI Questions
- Reminder to practice with equipment/materials and review the handbook prior to beginning work
- Review of Certification process; handout information, schedule
- Training evaluation and distribute training checks

This page intentionally left blank.

Appendix G: SCID Editing Form

2012 MENTAL HEALTH SURVEILLANCE STUDY (MHSS)

SCID EDITING FORM

CI Name: _____ Type of Review _____Full _____ Part
Quest ID: _____ CS Name _____
Date of Interview: _____ Date of review/edit: _____
Feedback given by: _____ Date of Feedback: _____

Summary of case (demographics, dx's, salient points, GAF)

_____

_____

_____

√= ok/went into section + = very good – = see comments below

OV	MDE-py	MDE-lt	Manic – py	Manic – lt	Dysthymia	Psychotic	Mood Diff	PTSD	Panic	Social	Specific	OCD	GAD	Anorexia	Bulimia	IED	Alcohol	Drugs	Adjustment	GAF	End of Interview

Adequate notes	Missed codes	Proper codes	Proper probes	Rapport w R	Dealt w/ emotions properly	Asked question verbatim	Got description in R's own words	Behavioral referents	Right amt of information	Resolved contradictions	Medical or substance etiologies	Correct time period	Differentiated bet symptoms easily confused	Leading questions	Distressed respondent	Completed scid / followed skips

Note both strengths and areas to develop. Use examples from interview.

KEY
Adequate notes = provided adequate notes to support ratings; Adjustment = Adjustment Disorder Module; Alcohol = Alcohol Use Disorders Module; Anorexia = Anorexia Nervosa Module; Asked question verbatim = asked the initial questions verbatim; Behavioral referents = obtained examples to support endorsed symptoms; Bulimia = Bulimia Nervosa Module; Completed scid / followed skips = properly administered the SCID; Correct time period = anchored the respondent in the correct time periods; Dealt w/ emotions properly = properly responded to respondent's emotions; Differentiated symptoms easily confused = properly differentiated symptoms that overlap; Distressed respondent = properly followed the Distressed Respondent Protocol; Drugs = Non-Alcohol Substance Use Disorders Module; Dysthymia = Dysthymic Disorder Module; End of interview = read End of Interview script (K.4); GAD = Generalized Anxiety Disorder Module; GAF = Global Assessment of Functioning Scale; Got description in R's own words =

documented symptoms in the respondent's own words; IED = Intermittent Explosive Disorder Module; Leading questions = did not ask leading questions; Manic – lt = Manic Episode-Lifetime Module; Manic – py = Manic Episode-Past Year Module; MDE-lt = Major Depressive Episode-Lifetime Module; MDE-py = Major Depressive Episode-Past Year Module; Medical or substance etiologies = ruled out disorders due to general medical conditions, substance use, or medication; Missed codes = did not circle all codes; Mood Diff = Mood Differential Module; OCD = Obsessive-Compulsive Disorder Module; OV = Overview Module; Panic = Panic Disorder Module; Proper codes = properly coded the data; Proper probes = asked proper follow-up questions; Psychotic = Psychotic Screening Module; PTSD = Posttraumatic Stress Disorder Module; Rapport w R = established and maintained rapport with respondent; Resolved contradictions = asked follow-up probes to resolve conflicting information; Right amt of information = obtained the right amount of information to have confidence in data; Social = Social Phobia Module; Specific = Specific Phobia Module.

Appendix H: SCID Technical Editing

SCID Technical Editing Guide……..H-1
SCID Failed Edit Worksheet…….H-15

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

Mental Health Surveillance Study

SCID Technical Editing Guide

January 2012

2012 MENTAL HEALTH SURVEILLANCE STUDY (MHSS)

SCID Technical Edit Guide

Technical Editors should check all variables as specified. should be notified of any inconsistencies or cases in which conditions are not true.

When Technical Editing is complete, a code 89 should be entered into CMS system (https://nsduhweb.rti.org/nsduhFT/login.cfm) and your name and the date should be recorded on the cover of the SCID booklet in the fields marked "Edited by:" and "Date Edited"

A. CHECK FRONT COVER

1. On front cover of SCID booklet, check that all information is correctly filled in:

Var Name	# of Characters	Range	Description
Interviewer ID:	6	000998 – 680447	If blank or not the right number of characters, this can be found in the CMS and should be filled in correctly
QuestID:	7	2000000 – 9969999	If blank or not the right number of characters, contact to determine what the correct QUESTID is.
Date of Interview:	6	MM/DD/YY	If blank, this can be found in the CMS (it is the date of event 70, 71, or 72) and should be filled in
Date Shipped to RTI:	6	MM/DD/YY	If blank, this can be found in the CMS (it is the date of event 80) and should be filled in
Date Received at RTI:	6	MM/DD/YY	If blank, this can be found in the CMS (it is the date of event 81) and should be filled in
Clinical QC by:	20	Unrestricted	This should be the name or initials of the Clinical Supervisor who reviewed the SCID. If blank, this can be found in the CMS (it is the name associated with events 86 and 87) and should be filled in
Date of Clinical QC:	6	MM/DD/YY	If blank, this can be found in the CMS (it is the date of event 87) and should be filled in
Edited by:	20	Unrestricted	This should be filled in with the technical editor's name
Date Edited:	6	MM/DD/YY	This should be filled in with the date the Technical Edit is completed.

B. CHECK REQUIRED FIELDS

1. Check that all required variables have been filled.

Var Name	Range	Description
ISTH	0–12	Interview Start Time – Overview Hour
ISTM	00–59	Interview Start Time – Overview Minute
AMPM1	AM, PM	
OV1	1, 2, 8	Respondent's sex
OV2	MM/DD/YYYY, 88/88/8888	Respondent's date of birth
OV3	1–5, 8	Marital status
OV4	1–8, 88	Education (highest level completed)
OV5	1, 3, 8	Treatment for emotions [IF OV5=1, SKIP to OV6]
OV5a	1-4, 8	History of mental health treatment
OV6	0–5, 8	Number of hospitalizations [IF OV6=0, SKIP to OV7a]
OV6a	1-2, 8	History of hospitalizations [IF OV6a≠2 skip to OV7a]
OV6b	Open text	Reason(s) for Psychiatric Hospitalization(s) - Past Year
OV7a	Open text	Psychotropic medications taken in the past year – NOT currently
OV7b	Open text	Psychotropic medications taken in the past year - currently
EOI1	0–100	Global Assessment of Functioning (GAF) Scale; 0= inadequate information **(located at end of SCID behind End of Interview tab).**
IDS1	1, 3, 8	Distressed Respondent Protocol used **(located at end of SCID behind Interview Debriefing tab)**
IDS3	1, 3, 8	Short Blessed Scale used [If IDS3 = 1, SKIP to IDS5]
IDS4a	0-28, 88	Score on Short Blessed
IDS5	1, 3, 8	Primary support group problems
IDS6	1, 3, 8	Social environment problems
IDS7	1, 3, 8	Educational problems
IDS8	1, 3, 8	Occupational problems
IDS9	1, 3, 8	Housing problems
IDS10	1, 3, 8	Economic problems
IDS11	1, 3, 8	Health care access problems
IDS12	1, 3, 8	Criminal justice problems
IDS13	1, 3, 8	Life-threatening illness – self
IDS14	1, 3, 8	Life-threatening illness – family
IDS15	1, 3, 8	Other psychosocial/environment problems
IDS16	1–5, 8	Comprehension Rating
IDS18	1–5, 8	Cooperation Rating
IDS20	1–5, 8	Privacy Rating
IDS22	1–5, 8	Global Validity Rating
IDS24	1, 3, 8	Potential disorders not assessed [IF IDS24 = 1; SKIP to IDS26]
IDS24a	1-4, 8	Rule-out disorder present
IDS26	1–5, 8	CS: Global Validity Rating
IDS28	1, 3, 8	CS: Potential disorders not assessed [IF IDS28=1; SKIP IDS28a and IDS29]
IDS28a	1-4, 8	Rule-out disorder present

C. CHECK SCREENING VARIABLES FOR CONSISTENCY

1. On "Screening-Page 1" and "Screening-Page 2" check variables S1-S9:

Var Name	Range	Description	Consistency Check
S1	1–3, 8	Panic attacks in past year	if S1=1 then E28=1; if S1=2 or 3 then E28=3

S2	1–3, 8	Fear of leaving the house, being alone in past year	if E52=3 or BLANK then E53=Blank (ignore S2) else;
			if S2=1 then E53=1;
			if S2=2 or 3 then E53=3

S3	1–3, 8	Fear of being around others in past year	if S3=1 then E60=1; if S3=2 or 3 then E60=3
S4	1–3, 8	Fear of flying, blood, being enclosed in past year	if S4=1 then E70=1; if S4=2 or 3 then E70=3
S5	1–3, 8	Bothered by illogical/repeated thoughts in past year	if S5=1 then E78=1; if S5=2 or 3 then E78=3
S6	1–3, 8	Compulsive behavior in past year	if S6=1 then E83=1; if S6=2 or 3 then E83=3
S7	1–3, 8	Nervous/anxious Past Year	if S7=1 then E93=1; if S7=2 or 3 then E93=3
S8	1–3, 8	Underweight	if S8=1 then F1=1; if S8=2 or 3 then F1=3
S9	1–3, 8	Out of control eating	If F7=3 then F8=blank, else if S9=1 then F8=1; if S9=2 or 3 then F8=3

D. CHECK SKIP PATTERNS

For these variables check to see if the conditions stated in the "Skip Logic" column are true. For example, if variable A2 has a 1 circled, then variables A3 through A17 should be blank. If they are not all blank, then write down the name of the variable on the failed edit worksheet that has a response entered and the variable from which the skip originated (listed in the first column tables that follow). So if for example, A2 = 1 and you look through an notice A5 also has a 1 circled then you should write the following in the columns. You do not have to write anything in the notes column. Example:

Variable Name (Skip begins)	Var Name (Should be blank)	Notes
A2=1	A5=1	

1. MODULE A: MOOD EPISODES

Var Name	Range	Description	Skip Logic
A2	1–3, 8	A2- Loss of interest or pleasure	if A1≠ 3 and A2≠ 3 then A3–A17=BLANK
A10	1, 3, 8, 9	Criteria A met	if A10=1 then A11–A17=BLANK
A11	1–3, 8, 9	Clinically significant distress or impairment	if A11=1 or 2 then A12–A17=BLANK
A14	1, 3, 8, 9	Primary Mood Episode (not due to substance or GMC)	if A14=1 then A15–A17=BLANK
A15	1, 3, 8, 9	At least 1 episode not simple bereavement	if A15=1 then A16 and A17=BLANK

Var Name	Range	Description	Skip Logic
A16	1, 3, 8, 9	Past Year Major Depressive Episode	if A16=1 then A17=BLANK; if A16=3 then A17=2 digit number and A18–A34=BLANK
A19	1–3, 8, 9	A2 - Loss of interest or pleasure	if A18≠ 3 and if A19≠ 3 then A20–A34=BLANK
A27	1, 3, 8, 9	Criteria A met	if A27=1 then A28–A34=BLANK
A28	1–3, 8, 9	Clinically significant distress or impairment	if A28=1 or 2 then A29–A34=BLANK
A30	1, 3, 8, 9	Primary Mood Episode (not due to substance or GMC)	if A30=1 then A31–A34=BLANK
A31	1, 3, 8, 9	At least 1 Episode not Simple bereavement	if A31=1 then A32–A34=BLANK
A32	1, 3, 8, 9	Lifetime Major Depressive Episode	if A32=1 then A34=BLANK
A35	1–3, 8	A1 - Abnormally elevated, expansive or irritable mood	if A35=1 then A36–A50=BLANK
A36	1–3, 8	A2 – Irritability/elevated mood at least 1 week	if A36=1 or 2 then A37–A50=BLANK
A44	1, 3, 8, 9	Criteria B met	If A44=1 then A45–A50=BLANK
A45	1–3, 8, 9	Criteria D met	If A45=1 or 2 then A46–A50=BLANK
A48	1, 3, 8, 9	Criteria E met	If A48=1 then A49 and A50=BLANK
A49	1, 3, 8, 9	Manic Episode past year	If A49=1 then A50=BLANK; if A49=3 then A50 has a response and A51–A80=BLANK
A51	1–3, 8, 9	Criteria A	If A51=1 then A52–A65=BLANK
A52	1–3, 8, 9	Criteria A2 – lasting one week or more	If A52=1 or 2 then A53–A65=BLANK
A60	1, 3, 8, 9	Criteria B	If A60=1 then A61–A65=BLANK
A61	1–3, 8, 9	Criteria D	If A61=1 or 2 then A62–A65=BLANK
A63	1, 3, 8, 9	Criteria E- Primary Mood	If A63=1 then A64–A65=BLANK
A64	1, 3, 8, 9	Manic Episode criteria met	If A64=1 then A65=BLANK; if A64=3 then A65 has a response and A66–A80=BLANK
A66	1–3, 8, 9	Criteria A met	If A66=1 then A67–A80=BLANK
A73	1–3, 8, 9	Criteria B met	If A73=1 then A74–A80=BLANK
A74	1, 3, 8, 9	Criteria C met	If A74=1 then A75–A80=BLANK
A75	1, 3, 8, 9	Criteria D met	If A75=1 then A76–A80=BLANK
A77	1, 3, 8, 9	Criteria F met	If A77=1 then A78–A80=BLANK
A78	1, 3, 8, 9	Criteria G met	If A78=1 then A79 and A80=BLANK
A79	1–3, 8, 9	Criteria H - Clinically significant distress or impairment	If A79=1 then A80=BLANK

2. MODULE B/C: PSYCHOTIC SCREENING

Var Name	Range	Description	Skip Logic
BC1	1–3, 8	Delusion of Reference	if BC1=1 or 2 then BC2=BLANK
BC3	1–3, 8	Persecutory Delusion	if BC3=1 or 2 then BC4=BLANK
BC5	1–3, 8	Grandiose Delusion	if BC5=1 or 2 then BC6=BLANK
BC7	1–3, 8	Somatic Delusion	if BC7=1 or 2 then BC8=BLANK
BC9	1–3, 8	Other Delusions	if BC9=1 or 2 then BC10=BLANK
BC11	1–3, 8	Auditory Hallucinations	if BC11=1 or 2 then BC12–BC14=BLANK
BC15	1–3, 8	Visual Hallucinations	if BC15=1 or 2 then BC16=BLANK
BC17	1–3, 8	Tactile Hallucinations	if BC17=1 or 2 then BC18=BLANK
BC19	1–3, 8	Other hallucinations, e.g. gustatory, olfactory	if BC19=1 or 2 then BC20=BLANK

3. MODULE D: MOOD DISORDERS

Var Name	Range	Description	Skip Logic
D1	1, 3, 8	No Clinically Significant Mood SXS	if D1=1 then D2–D9=BLANK
D2	1, 3, 8, 9	History of 1+ Manic or Mixed Episodes	if D2=1 then D3–D5=BLANK
D3	1, 3, 8, 9	At least 1 Manic/Mixed Episode Not due to GMC/Substance Use	if D3=1 then D4–D5=BLANK
D4	1, 3, 8, 9	Bipolar I Disorder	if D4=1 then D5=BLANK; if D4=3 then D5 is filled in and D6–D9=BLANK
D6	1, 3, 8, 9	At least 1 MDE Not due to GMC/Substance Use	if D6=1 then D7–D9=BLANK
D7	1, 3, 8, 9	At least 1 MDE Not accounted for by Schizoaffective Disorder	if D7=1 then D8 and D9=BLANK
D8	1, 3, 8, 9	Major Depressive Disorder	if D8=1 then D9=BLANK

4. MODULE E: ANXIETY DISORDERS

a. Post Traumatic Stress Disorder

Var Name	Range	Description	Skip Logic
E1	1, 3, 8	Traumatic Events	if E1=1 then E2–E27=BLANK
E2	1, 3, 8, 9	Presence of Upsetting Memories	if E2=1 then E3–E27=BLANK
E3	1–3, 8, 9	A1 - Event involved actual/threatened death or serious injury	if E3=1 then E4–E27=BLANK
E4	1–3, 8, 9	A2 - Response involved fear/helplessness, horror	if E4=1 then E5–E27=BLANK
E10	1, 3, 8, 9	Criteria B met	if E10=1 then E11–E27=BLANK
E18	1, 3, 8, 9	Criteria C met	if E18=1 then E19–E27=BLANK
E24	1, 3, 8, 9	Criteria D met	if E24=1 then E25–E27=BLANK
E25	1–3, 8, 9	E - Duration of disturbance > 1 month	if E25=1 then E26 and E27=BLANK
E26	1–3, 8, 9	F - Clinically significant distress/impairment	if E26=1 then E27=BLANK

b. Panic Disorder

Var Name	Range	Description	Skip Logic
E28	1, 3, 8	S1 - Panic Disorder - Screening Question 1	if E28=1 then E29–E51=BLANK
E29	1–3, 8, 9	A1 - Recurrent/unexpected panic attacks	if E29=1 then E30–E51=BLANK
E30	1–3, 8, 9	A2- one month + concern about panic attacks	if E30=1 then E31–E51=BLANK
E31	1–3, 8, 9	Panic attack symptoms developed abruptly	if E31=1 then E32–E51=BLANK
E45	1, 3, 8, 9	Qualified Panic Attack - At least 4 symptoms and E31=3	if E45=1 then E46–E51=BLANK
E46	1, 3, 8, 9	C - Primary Anxiety Disorder	if E46=1 then E47–E51=BLANK
E47	1, 3, 8, 9	D – Panic Attacks not due to other mental disorder	
E47a	1, 3, 8, 9	Panic Disorder	If E47a=1 then E48-E51=BLANK
E48	1–3, 8, 9	B1 - Presence of Agoraphobia	if E48=1 then E49–E59=BLANK
E49	1–3, 8, 9	B2 - Avoidance of Agoraphobic situations	if E49=1 then E50–E59=BLANK
E50	1, 3, 8, 9	B3- Agoraphobia Not due to other disorder	if E50=1 then E51–E59=BLANK

c. Agoraphobia without Panic Disorder (AWOPD)

Var Name	Range	Description	Skip Logic
E52	1, 3, 8, 9	Past Year Panic Disorder Criteria Met	if E52=3 then E53–E59=BLANK
E53	1, 3, 8, 9	S2 – AWOPD – Screening Question 2	if E53=1 then E54–E59=BLANK
E54	1–3, 8, 9	A1 - Presence of Agoraphobia	if E54=1 then E55–E59=BLANK
E55	1–3, 8, 9	A2 - Agoraphobic situations avoided	if E55=1 then E56–E59=BLANK
E56	1, 3, 8, 9	A3 - Avoidance not due to other mental disorder	if E56=1 then E57–E59=BLANK
E57	1, 3, 8, 9	Criteria A met	if E57=1 then E58 and E59=BLANK
E58	1, 3, 8, 9	C - Primary Anxiety Disorder (not due to GMC/Substance)	if E58=1 then E59=BLANK

d. Social Phobia

Var Name	Range	Description	Skip Logic
E60	1, 3, 8	S3 – Social Phobia – Screening Question 3	if E60=1 then E61–E69=BLANK
E61	1–3, 8, 9	A – Fear of social performance situation	If E61=1 then E62–E69=BLANK
E62	1–3, 8, 9	B - Exposure to social situation provoke anxiety	if E62=1 then E63–E69=BLANK
E63	1–3, 8, 9	C- Recognize excessive fear	if E63=1 then E64–E69=BLANK
E64	1–3, 8, 9	D - Avoid social or performance situations	if E64=1 then E65–E69=BLANK
E65	1–3, 8, 9	E - Avoidance/anxiety interfere with functioning	if E65=1 then E66–E69=BLANK
E66	1, 3, 8, 9	G- Primary Anxiety Disorder (not due to substance/GMC)	if E66=1 then E67–E69=BLANK

E67	1–3, 8, 9	Phobia not due to other disorder	if E67=1 then E68 and E69=BLANK
E68	1–3, 8, 9	H - Phobia unrelated to GMC	if E68=1 then E69=BLANK

e. Specific Phobia

Var Name	Range	Description	Skip Logic
E70	1, 3, 8	S4 – Specific Phobia – Screening Question 4	if E70=1 then E71–E77=BLANK
E71	1–3, 8, 9	A - Excessive/unreasonable fear	if E71=1 then E72–E77=BLANK
E72	1–3, 8, 9	B - Exposure to phobic stimulus induce anxiety	if E72=1 then E73–E77=BLANK
E73	1–3, 8, 9	Criteria B met	if E73=1 then E74–E77=BLANK
E74	1–3, 8, 9	D - Avoidance of phobic situation (Criteria D)	if E74=1 then E75–E77=BLANK
E75	1–3, 8, 9	E - Avoidance/anxiety interfere with functioning	if E75=1 then E76 and E77=BLANK
E76	1, 3, 8, 9	G - Symptoms not due to other disorder	if E76=1 then E77=BLANK

f. Obsessive Compulsive Disorders (OCD)

Var Name	Range	Description	Skip Logic
E78	1, 3, 8	S5 – Obsessions – Screening Question 5	if E78=1 then E79–E82=BLANK
E79	1–3, 8, 9	A1 - Intrusive thoughts, images, impulses causing anxiety	if E79=1 then E80–E82=BLANK
E80	1–3, 8, 9	A2 - Intrusive thoughts are not excessive worries	if E80=1 then E81 and E82=BLANK
E81	1–3, 8, 9	A3 - Ignore/repress intrusive thoughts. Images, impulses	if E81=1 then E82=BLANK
E83	1, 3, 8	S6 – Compulsions – Screening Question 6	if E83=1 then E84 and E85=BLANK
E86	1, 3, 8	Presence of obsession or compulsion	if E86=1 then E87–E92=BLANK
E87	1–3, 8, 9	B - Recognize obsessions/compulsions excessive	if E87=1 then E88–E92=BLANK
E89	1–3, 8, 9	C - Obsessions/compulsions interfere with normal functioning	if E89=1 then E90–E92=BLANK
E90	1, 3, 8, 9	D - Obsessions/compulsions not related to Axis I Disorder	if E90=1 then E91 and E92=BLANK
E91	1, 3, 8, 9	E - Primary Anxiety Disorder (not due to Subst/GMC)	if E91=1 and E92=BLANK

g. Generalized Anxiety Disorder (GAD)

Var Name	Range	Description	Skip Logic
E93	1, 3, 8	S7 – GAD – Screening Question 7	If E93=1 then E94–E107=BLANK
E94	1–3, 8, 9	A - Excessive anxiety past 6 months	if E94=1 then E95–E107=BLANK
E95	1–3, 8, 9	B - Difficult to control worry past 6 months	if E95=1 then E96–E107=BLANK
E96	1, 3, 8, 9	Does Not Co-occur with Mood, psychotic disorder	if E96=1 then E97–E107=BLANK
E103	1–3, 8, 9	Criteria C met	if E103=1 then E104–E107=BLANK
E104	1, 3, 8, 9	D - SX Not related to Axis 1 disorder	if E104=1 then E105–E107=BLANK
E105	1–3, 8, 9	E - Clinically significant distress/impairment	if E105=1 then E106 and E107=BLANK
E106	1, 3, 8, 9	F - Primary Anxiety Disorder (not due to Subst/GMC)	if E106=1 then E107=BLANK

5. MODULE F: EATING DISORDERS

Var Name	Range	Description	Skip Logic
F1	1, 3, 8	S8 – Anorexia Nervosa – Screening Question 8	if F1=1 then F2–F6=BLANK
F2	1–3, 8, 9	A - Low body weight	if F2=1 then F3–F6=BLANK
F3	1–3, 8, 9	B - Fear of gaining weight while underweight	if F3=1 then F4–F6=BLANK
F4	1–3, 8, 9	C - Distorted body image	if F4=1 then F5 and F6=BLANK
F5	1–3, 8, 9	D – Amenorrhea (if OV=1 (Male R) F5 will be blank	if F5=1 then F6=BLANK
F7	1, 3, 8	Criteria Met for Anorexia	if F7=3 then F8–F15=BLANK
F8	1, 3, 8, 9	S9 – Bulimia Nervosa – Screening Question 9	if F8=1 then F9–F15=BLANK
F9	1–3, 8, 9	A2 - No control over eating	if F9=1 then F10–F15=BLANK
F10	1–3, 8, 9	A1 - Binge eating	if F10=1 then F11–F15=BLANK
F11	1–3, 8, 9	B - Behavior to prevent weight gain (vomit, laxative)	if F11=1 then F12–F15=BLANK
F12	1–3, 8, 9	C - Binge eating & compensatory behavior bi-weekly past 3 month	if F12=1 then F13–F15=BLANK
F13	1–3, 8, 9	D - Self-evaluation based on body image	if F13=1 then F14 and F15=BLANK
F14	1, 3, 8, 9	E - Does not Co-occur with Anorexia	if F14=1 then F15=BLANK

6. MODULE G: IMPULSE CONTROL DISORDERS

Var Name	Range	Description	Skip Logic
G1	1–3, 8	A - Serious Assaultive acts	if G1=1 then G2–G4=BLANK
G2	1–3, 8, 9	B - Extreme aggressiveness	if G2=1 then G3 and G4=BLANK
G3	1–3, 8, 9	C - SX due to other disorder	if G3=1 then G4=BLANK

7. MODULE H: SUBSTANCE USE DISORDERS

Var Name	Range	Description	Skip Logic
H1	1, 3, 8	Drank 6 + drinks in Past year	if H1=1 then H2–H14=BLANK
H15	1, 3, 8	Sedatives - Hypnotics- Anxiolytics	if H15=1 then H24, H32, H40, H48, H56, H64, H72, H80, H88, H96, H104, H112, H120=BLANK
H16	1, 3, 8	Stimulants	if H16=1 then H25, H33, H41, H49, H57, H65, H73, H81, H89, H97, H105, H113, H121=BLANK
H17	1, 3, 8	Opioids	if H17=1 then H26, H34, H42, H50, H58, H66, H74, H82, H90, H98, H106, H114, H122=BLANK
H18	1, 3, 8	Cannabis	if H18=1 then H27, H35, H43, H51, H59, H67, H75, H83, H91, H99, H107, H115, H123=BLANK
H19	1, 3, 8	Heroin	if H19=1, then H28, H36, H44, H52, H60, H68, H76, H84, H92, H100, H108, H116, H124=BLANK
H20	1, 3, 8	Cocaine	if H20=1, then H29, H37, H45, H53, H61, H69, H77, H85, H93, H101, H109, H117, H125=BLANK

Var Name	Range	Description	Skip Logic
H21	1, 3, 8	Hallucinogens/PCP	if H21=1, then H30, H38, H46, H54, H62, H70, H78, H86, H94, H102, H110, H118, H126=BLANK
H22	1, 3, 8	Inhalants	if H22=1, then H31, H39, H47, H55, H63, H71, H79, H87, H95, H103, H111, H119, H127=BLANK
H23	1, 3, 8	Past Year Drug Use	if H23=1 then H24–H127=BLANK

8. MODULE J: ADJUSTMENT DISORDER

Var Name	Range	Description	Skip Logic
J1	1, 3, 8	Disturbance in Past Year not meeting other Axis 1 criteria	if J1=1 then J2–J8=BLANK
J2	1–3, 8, 9	A - Symptoms due to identifiable stressor within 3 months of onset	if J2=1 then J3–J8=BLANK
J3	1–3, 8, 9	B - Clinically significant distress/impairment	if J3=1 then J4–J8=BLANK
J4	1, 3, 8, 9	C- Not due to other Axis I or II disorder	if J4=1 then J5–J8=BLANK
J5	1, 3, 8, 9	D - Symptoms not due to Bereavement	if J5=1 then J6–J8=BLANK
J6	1–3, 8, 9	E - Symptoms do not persist for 6+ months	if J6=1 then J7 and J8=BLANK
J7	1, 3, 8, 9	Adjustment Disorder	if J7=1 then J8=BLANK

E. CHECK SYMPTOM CRITERIA & DIAGNOSITIC VARIABLES TO ENSURE CONDITIONS MET FOR CODE 3

1. MODULE A: MOOD EPISODES

a. Past Year and Lifetime Major Depressive Episode

Var Name	Range	Description	Conditions
A10	1, 3, 8, 9	Symptom A Criteria Pat Year MDE	A10=3 if at least 5 of A1–A9=3 and (A1=3 or A2=3)
A16	1, 3, 8, 9	Past Year Major Depressive Episode	A16=3 if A10=3 and A11=3 and A14=3 and A15=3
A27	1, 3, 8, 9	Symptom A – Lifetime MDE	A27=3 if at least 5 of A18–A26=3 and (A18=3 or A19=3)
A32	1, 3, 8, 9	Lifetime Major Depressive Episode	A32=3 if A27=3 and A28=3 and A30=3 and A31=3

b. Past Year and Lifetime Manic Episode

Var Name	Range	Description	Conditions
A44	1, 3, 8, 9	Criteria B – Manic Past Year met if manic mood, then at least 3 of A37-A43=3; if only irritable mood (A44a=3), then 4 of A37-A43=3.	A44=1 if fewer than 3 of A37–A43=3 or if 3 of A37–A43=3 and A44a=3
A49	1, 3, 8, 9	Criteria A, B, D, E Manic Past Year	A49=3 if A35=3 and A36=3 and A44=3 and A45=3 and A48=3
A60	1, 3, 8, 9	Criteria B - Lifetime Manic - met if manic mood, then at least 3 of A53-A59= 3; if only irritable mood (A60a=3), then 4 of A53-59 = 3.	A60=1 if fewer than 3 of A53–A59=3 or if 3 of A53–A59=3 and A60a=3
A64	1, 3, 8, 9	Lifetime Manic Episode	A64=3 if A51=3 and A52=3 and A60=3 and A61=3 and A63=3

c. Dysthymic Disorder

Var Name	Range	Description	Conditions
A73	1–3, 8, 9	Criteria B met	A73=3 if at least 2 of A67–A72=3
A80	1, 3, 8, 9	Dysthymic Disorder	A80=3 if A66=3 and A73=3 and A74=3 and A75=3 and A77=3 and A78=3 and A79=3

2. MODULE B/C: PSYCHOTIC SCREENING

Var Name	Range	Description	Conditions
BC21	1, 3, 8	Primary Psychotic symptom present	BC21=3 if (BC1=3 and BC2=3) or (BC3=3 and BC4=3) or (BC5=3 and BC6=3) or (BC7=3 and BC8=3) or (BC9=3 and BC10=3) or (BC11=3 and BC12=3) or (BC15=3 and BC16=3) or (BC17=3 and BC18=3) or (BC19=3 and BC20=3)

3. MODULE E: ANXIETY DISORDERS

Var Name	Range	Description	Conditions
E10	1, 3, 8, 9	Criteria B PTSD	E10=3 if at least 1 of E5–E9=3
E18	1, 3, 8, 9	Criteria C PTSD	E18=3 if at least 3 of E11–E17=3
E24	1, 3, 8, 9	Criteria D PTSD	E24=3 if at least 2 of E19–E23=3
E27	1, 3, 8, 9	Post-traumatic stress disorder in past year	E27=3 if E3=3 and E4=3 and E10=3 and E18=3 and E24=3 and E25=3 and E26=3
E45	1, 3, 8, 9	Qualified Panic attack	E45=3 if at least 4 of E32–E44=3 and E31=3
E47a	1, 3, 8, 9	Panic Disorder	E47a=3 if E29=3 and E30=3 and E45=3 and E46=3 and E47=3

E51	1, 3, 8, 9	Panic Disorder with Agoraphobia in past year	E51=3 if E48=3 and E49=3 and E50=3

Var Name	Range	Description	Conditions
E52	1, 3, 8, 9	Past Year Panic Disorder Criteria Met	E52=3 if E51=1 or 2 or 3
E57	1, 3, 8, 9	Criteria A met	E57=3 if E54=3 and E55=3 and E56=3
E69	1, 3, 8, 9	Social Phobia in past year	E69=3 if E61=3 and E62=3 and E63=3 and E64=3 and E65=3 and E66=3 and E67=3 and E68=3
E77	1, 3, 8, 9	Specific Phobia Past Year	E77=3 if E71= 3 and E72=3 and E73=3 and E74=3 and E75=3 and E76=3
E103	1–3, 8, 9	Criteria C met	E103=3 if at least 3 of E97–E102=3
E107	1, 3, 8, 9	Generalized Anxiety Disorder	E107=3 if E94=3 and E95=3 and E96=3 and E103=3 and E104=3 and E105=3 and E106=3

4. MODULE F: EATING DISORDERS

Var Name	Range	Description	Conditions
F6	1, 3, 8, 9	Anorexia Nervosa Past Year	F6=3 if F2=3 and F3=3 and F4=3 and (F5=3 or F5=Blank and OV1=1)
F15	1, 3, 8, 9	Bulimia Nervosa Past Year	F15=3 if F9=3 and F10=3 and F11=3 and F12=3 and F13=3 and F14=3

5. MODULE G: IMPULSE CONTROL DISORDERS

Var Name	Range	Description	Conditions
G4	1, 3, 8, 9	Intermittent Explosive Disorder Past year	G4=3 if G1=3 and G2=3 and G3=3

6. MODULE H: SUBSTANCE USE DISORDERS

Var Name	Range	Description	Conditions
H9	1, 3, 8, 9	Alcohol Dependence	H9=3 if at least 3 of H2–H8=3
H14	1, 3, 8, 9	Alcohol Abuse	H14=3 if at least 1 of H10–H13=3
H23	1, 3, 8, 9	Past Year Threshold Substance Use	H23=3 if at least one of H15–H22=3
H80	1, 3, 8, 9	Sedatives - hypnotics- anxiolytics	H80=3 if at least 3 of H24, H32, H40, H48, H56, H64, H72=3
H81	1, 3, 8, 9	Stimulants	H81=3 if at least 3 of H25, H33, H41, H49, H57, H65, H73=3
H82	1, 3, 8, 9	Opioids	H82=3 if at least 3 of H26, H34, H42, H50, H58, H66, H74=3
H83	1, 3, 8, 9	Cannabis	H83=3 if at least 3 of H27, H35, H43, H51, H59, H67, H75=3

H84	1, 3, 8, 9	Heroin	H84=3 if at least 3 of H28, H36, H44, H52, H60, H68, H76=3
H85	1, 3, 8, 9	Cocaine	H85=3 if at least 3 of H29, H37, H45, H53, H61, H69, H77=3
Var Name	**Range**	**Description**	**Conditions**
H86	1, 3, 8, 9	Hallucinogens/PCP	H86=3 if at least 3 of H30, H38, H46, H54, H62, H70, H78=3
H87	1, 3, 8, 9	Inhalants	H87=3 if at least 3 of H31, H37, H47, H55, H63, H71, H79=3
H120	1, 3, 8, 9	Sedatives - Hypnotics- Anxiolytics	H120=3 if at least 1 of H88, H96, H104, H112=3
H121	1, 3, 8, 9	Stimulants	H121=3 if at least 1 of H89, H97, H105, H113=3
H122	1, 3, 8, 9	Opioids	H122=3 if at least 1 of H90, H98, H106, H114=3
H123	1, 3, 8, 9	Cannabis	H123=3 if at least 1 of H91, H99, H107, H115=3
H124	1, 3, 8, 9	Heroin	H124=3 if at least 1 of H92, H100, H108, H116=3
H125	1, 3, 8, 9	Cocaine	H125=3 if at least 1 of H93, H101, H109, H117=3
H126	1, 3, 8, 9	Hallucinogens/PCP	H126=3 if at least 1 of H94, H102, H110, H118=3
H127	1, 3, 8, 9	Inhalants	H127=3 if at least 1 of H95, H103, H111, H119=3

7. MODULE J. ADJUSTMENT DISORDER

Var Name	Range	Description	Conditions
J7	1, 3, 8, 9	Adjustment Disorder	J7=3 if J2=3 and J3=3 and J4=3 and J5=3 and J6=3

This page has been intentionally left blank.

Mental Health Surveillance Study (MHSS)

SCID Failed Edit Worksheet

Quest ID:_____ Technical Editor: _____ Date: _____

Clinical Interviewer: _____ Clinical Supervisor: _____

A. Check Front Cover

	Interviewer ID
	Quest ID
	Date of Interview
	Date Shipped to RTI

	Date Received at RTI
	Clinical QC by
	Date of Clinical QC

B. Check Required Fields – Problem is missing or out of range

	OV1
	OV2
	OV3
	OV4
	OV5
	OV5a
	OV6
	OV6a
	OV6b
	OV7a
	OV7b
	EOI1 (End of interview, GAF)

	IDS1
	IDS3
	IDS4a
	IDS5
	IDS6
	IDS7
	IDS8
	IDS9
	IDS10
	IDS11
	IDS12
	IDS13
	IDS14

	IDS15
	IDS16
	IDS18
	IDS20
	IDS22
	IDS24
	IDS24a
	IDS26
	IDS28
	IDS28a

C. Required Fields & Check screening variables for consistency – problem is inconsistency

	S1
	S2
	S3

	S4
	S5
	S6

	S7
	S8
	S9

D. Check Skip Patterns - Legitimate Skip but data is present

Variable Name (Skip begins)	Variable Name (Should be blank, but response is circled. Include what response is written on SCID)	Notes

E. Check for symptom Criteria & Diagnostic Variables to ensure conditions met for code 3.

	A10
	A16
	A27
	A32
	A44
	A49
	A60
	A64
	A73
	A80

	BC21

	E10
	E18
	E24
	E27

	E45
	E47a
	E51
	E57
	E69
	E77
	E103
	E107

	F6
	F15

	G4

	H9
	H14
	H80
	H81

	H82
	H83
	H84
	H85
	H86
	H87
	H120
	H121
	H122
	H123
	H124
	H125
	H126
	H127

	J7

Appendix I: Clinical Interview Current Status Report

Total Number of Cases Not Keyed	112	23	120	24	100	19	118	22	450	22
	Quarter 1	%	Quarter 2	%	Quarter 3	%	Quarter 4	%	Total	%
Current Clinical Interview Status Codes										
00: Initialized -- Not worked	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
50: Appointment for Follow-Up Interview	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
51: R Unavailable	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
52: Ring No Answer, Busy	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
53: Machine/Voicemail: No Msg	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
54: Machine/Voicemail: Left Msg	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
56: Breakoff (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
57: Refusal	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
59: Other, Specify	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
70: Interview Completed, Audio Recorded	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
71: Interview Completed, No Audio (Refusal)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
72: Interview Completed, No Audio (Technical Problem)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
73: Breakoff (Partial Interview)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
74: Unable to Contact after Repeated Calls	77	15.49	75	15.03	64	11.99	81	14.92	297	14.33
75: Final - Phone Number Problem	4	0.80	17	3.41	6	1.12	6	1.10	33	1.59
76: Final - Refusal	3	0.60	4	0.80	2	0.37	3	0.55	12	0.58
77: Final - Other (Specify)	0	0.00	1	0.20	2	0.37	1	0.18	4	0.19
79: Case reworked - R not selected	3	0.60	3	0.60	0	0.00	0	0.00	6	0.29
80: Materials Shipment (SCID)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
81: SCID Received at Regent	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
82: SCID Received at Editing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
83: Not Ready for Editing - Audio file expected but not uploaded	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
84: Ready for Editing - No audio file captured (code 71 or 72)	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
85: Ready for Editing - Audio file uploaded	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
86: Editing in Progress	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
87: Editing Complete	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
88: SCID Delivered to Technical Editing	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
89: Technical Editing Complete	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
90: 90 SCID Not Keyed - Reason Unspecified	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
90.1: Respondent-Induced Break-off	9	1.81	16	3.21	9	1.69	7	1.29	41	1.98
90.2: Comprehension Problems - Language Barrier	0	0.00	0	0.00	1	0.19	0	0.00	1	0.05
90.3: Respondent Distress	4	0.80	1	0.20	2	0.37	4	0.74	11	0.53
90.4: Cognitive Impairment - Short Blessed administered and failed	2	0.40	3	0.60	4	0.75	9	1.66	18	0.87
90.5: Cognitive Impairment suspected but not captured by the short blessed - Short Blessed administered and passed, yet invalid for other reasons	5	1.01	0	0.00	2	0.37	2	0.37	9	0.43
90.6: Data of Questionable Validity (e.g., insufficient data, denial of symptoms, inconsistent report)	2	0.40	0	0.00	7	1.31	5	0.92	14	0.68
90.7: Interviewer error (e.g. Improper scoring of the Short Blessed, failure to follow protocol)	3	0.60	0	0.00	1	0.19	0	0.00	4	0.19
95: Keyed	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
96: Keying Verified	385	77.46	379	75.95	434	81.27	425	78.27	1623	78.29
Current Clinical Supervisor Review Codes										
40: Not Selected for Clinical Supervisor Full Audio Review	345	92.99	327	93.16	368	88.25	357	91.07	1397	91.25
41: Selected for Clinical Supervisor Full Audio Review - Pending	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
42: Clinical Supervisor Full Audio Review, In Progress	0	0.00	1	0.28	2	0.48	2	0.51	5	0.33
43: Clinical Supervisor Full Audio Review, Completed, no problems	19	5.12	16	4.56	36	8.63	23	5.87	94	6.14
44: Clinical Supervisor Full Audio Review, Completed with Minor Problems	6	1.62	7	1.99	11	2.64	10	2.55	34	2.22
45: Clinical Supervisor Full Audio Review, Completed with Major Problems	1	0.27	0	0.00	0	0.00	0	0.00	1	0.07
Total (Codes 40 through 45)	**371**	**100**	**351**	**100**	**417**	**100**	**392**	**100**	**1531**	**100**

2012 NATIONAL SURVEY ON DRUG USE AND HEALTH

METHODOLOGICAL RESOURCE BOOK SECTION 16a: 2012 MENTAL HEALTH SURVEILLANCE STUDY DESIGN AND ESTIMATION REPORT

Substance Abuse and Mental Health Services Administration
Center for Behavioral Health Statistics and Quality
Rockville, Maryland 20857

June 2014

2012 MENTAL HEALTH SURVEILLANCE STUDY DESIGN AND ESTIMATION REPORT

Prepared for the 2012 Methodological Resource Book (Section 16a)

Contract No. HHSS283201000003C
RTI Project No. 0212800.001.112.008.006
Deliverable No. 39

RTI Authors:

Dan Liao
Phillip Kott
Jeremy Aldworth
Feng Yu
Rhonda Karg
Bonnie Shook-Sa
Teresa Davis

Project Director:

David Hunter

SAMHSA Project Officer:

Peter Tice

SAMHSA Authors:

Jonaki Bose
Joseph Gfroerer
Sarra Hedden

For questions about this report, please email Peter.Tice@samhsa.hhs.gov.

Prepared for Substance Abuse and Mental Health Services Administration, Rockville, Maryland

Prepared by RTI International, Research Triangle Park, North Carolina

June 2014

Recommended Citation: Center for Behavioral Health Statistics and Quality. (2014). *2012 National Survey on Drug Use and Health: Methodological Resource Book (Section 16a, 2012 Mental Health Surveillance Study: Design and Estimation Report)*. Substance Abuse and Mental Health Services Administration, Rockville, MD.

Acknowledgments

This report would not be possible without the guidance and input of staff from the Center for Behavioral Health Statistics and Quality (CBHSQ) within the Substance Abuse and Mental Health Services Administration (SAMHSA). In particular, CBHSQ's Art Hughes provided useful comments. Special thanks are also due to several staff members at RTI International (a trade name of Research Triangle Institute). James R. Chromy, Katherine B. Morton, Rachel Harter, and Kevin Wang provided valuable comments on the documentation of the sample design, sample weighting, and estimation methods. Dave Hunter is the RTI Project Director. Debbie Bond and Valerie Garner word processed and formatted the report, while Richard Straw copyedited it and Teresa Bass coordinated its Web production.

Table of Contents

Chapter **Page**

List of Tables vii
List of Figures xi
List of Exhibits xi

1. Introduction 1

2. Instruments Used in Measuring Mental Illness 5
   2.1 Background 5
   2.2 Scales in the Main NSDUH Interview 5
      2.2.1 Psychological Distress Scale (K6) 5
      2.2.2 Functional Impairment Scales (WHODAS and SDS) 6
   2.3 Clinical Follow-Up Instruments 7

3. Sample Design and Selection 11
   3.1 Background 11
   3.2 Selecting the MHSS Clinical Follow-Up Samples 11
   3.3 2007 Feasibility Study 12
   3.4 2008 Sample Allocation 13
   3.5 2009 Sample Allocation 17
   3.6 2010, 2011, and 2012 Sample Allocations 20
   3.7 Creation of Variance Estimation Strata and Replicates 26
   3.8 Issues That Affected the Clinical Follow-Up Sample Design 29
      3.8.1 Assigning NSDUH Clinical Interview-Eligible Cases Zero Selection Probability 29
      3.8.2 Noncoverage for Non-English Speakers 30
      3.8.3 Gulf Coast Oversample 31
      3.8.4 Falsified Main NSDUH Interviews 31
      3.8.5 Respondents Excluded from 2008-2012 Adult Clinical Interview Data File 32

4. Data Collection and Processing 35
   4.1 Background 35
   4.2 SCID Data Collection 35
   4.3 SCID Data Processing 36
      4.3.1 Clinical Editing 37
      4.3.2 Technical Editing 38
      4.3.3 Keying 38
   4.4 Missing Data 38

5. Sample Weighting 43
   5.1 Background 43
   5.2 Components of the MHSS Analysis Weights 44
   5.3 Adjustment of Undercoverage of Hispanics 46
      5.3.1 Descriptive Characteristics of Hispanics Responding in English and Spanish 48

Table of Contents (continued)

Chapter **Page**

5.3.2 Development of the Coverage Adjustment to Compensate for Adults Interviewed in Spanish 49
5.4 Nonresponse Adjustment 51
5.4.1 Components of Nonresponse in the MHSS Clinical Follow-Up 51
5.4.2 Descriptive Characteristics of Adults Agreeing and Not Agreeing to Participate in the Clinical Follow-Up 54
5.4.3 Development of the Adjustment for Adults Not Agreeing to Participate in the Clinical Follow-Up 55
5.4.4 Descriptive Characteristics of Adults Completing and Not Completing the Clinical Follow-Up 58
5.4.5 Development of the Adjustment for Adults Not Completing the Clinical Follow-Up 59
5.4.6 Summary of Nonresponse Adjustments 62
5.5 Development of Revised Poststratification Adjustment and Scaling 62
5.5.1 Poststratification 62
5.5.2 Scaling 64
5.6 Evaluation of the Final Weights 66
5.6.1 Comparisons with the NSDUH Main Adult Sample 66
5.6.2 Unequal Weighting Factors 67
5.6.3 Comparisons Based on Standard Error Computations 68
5.7 Summary 70

6. MHSS Clinical Sample Characteristics 73
6.1 Background 73
6.2 Summary of Prior Sample Evaluations 73
6.3 Evaluating the Overall 2008 to 2012 Sample 76
6.3.1 Demographic Characteristics 77
6.3.2 Mental Health Characteristics 77
6.3.3 Substance Use Characteristics 78
6.3.4 K6 and WHODAS Total Scores 78
6.3.5 Summary 79

7. Estimation of Mental Illness for the 2008 to 2012 NSDUHs 91
7.1 Background 91
7.2 2008 Estimation Method 92
7.3 2012 Estimation Method 94
7.3.1 Investigating the 2008 (WHODAS) Model and Estimation Methods 94
7.3.2 Choosing Covariates in a Model for SMI 94
7.3.3 Metrics Used to Evaluate Models of SMI: Error Rate and Bias Measure 95
7.3.4 2012 WHODAS Model 98
7.3.5 2012 Model for the SDS Sample 99

Table of Contents (continued)

Chapter **Page**

7.4 Comparisons of SMI and AMI Estimates Using the 2012 and 2008 Estimation Methods 103
7.5 Caveats on Model-Based Mental Illness Prevalence Estimates 107

References 109

Appendix

A Kessler-6 (K6) Module A-1

B World Health Organization Disability Assessment Schedule (WHODAS) Module B-1

C Sheehan Disability Scale (SDS) Module C-1

D Descriptive Characteristics for Undercoverage of Hispanics, Agreement to Participate, and Completion of the Clinical Follow-Up D-1

E Descriptive Characteristics for Overall Impact of Weighting Adjustments E-1

List of Tables

Table **Page**

1.1 Mental Illness Categories Defined by SCID Disorder Diagnosis and GAF Score: 2008-2012 MHSS Clinical Follow-Up Study 3

2.1 Diagnostic Modules in the 2008-2012 MHSS SCID 8

3.1 2007 Mental Health Surveillance Feasibility Study Design Parameters 12

3.2 2008 Mental Health Surveillance Clinical Follow-Up Study Sample Allocation (*n* = 1,500) 15

3.3 Design Parameters for the 2008 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 16

3.4 2008 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary 17

3.5 2009 Mental Health Surveillance Clinical Follow-Up Study Sample Allocation: Modified Design (*n* = 500) 18

3.6 Design Parameters for the 2009 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 19

3.7 2009 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary 20

3.8 Age-Related Factors for the 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study 22

3.9 Projected Yields of Predicted Positive Cases, by Age Group: 2010 to 2012 MHSS Clinical Follow-Up 24

3.10 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study Sample Allocations, by K6 Group 24

3.11 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study Sample Allocations, by WHODAS Score 25

3.12 2010 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary 27

3.13 2011 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary 28

3.14 2012 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary 29

3.15 Completed NSDUH Main Interviews and MHSS Clinical Interviews in the GCO Region, by Year 31

3.16 Falsified Cases Removed from the 2008-2010 Mental Health Surveillance Clinical Follow-Up Study 32

List of Tables (continued)

Table **Page**

4.1 Unweighted Missing Numbers and Percentages of K6, WHODAS, and SDS Item Scores: 2008-2012 All-Adult NSDUH Main Study 40

4.2 Unweighted Missing Numbers and Percentages of MDE and Suicidal Thoughts: 2008-2012 All-Adult NSDUH Main Study 41

5.1 Adjusted Weights and Adjustment Factors in the NSDUH 2008-2012 Adult Clinical Interview Data File 47

5.2 Hispanic Respondents, by Language Version: 2008-2011 All-Adult NSDUH Main Study 48

5.3 Summary of Coverage Adjustments for Excluding Spanish-Language Version Respondents, by Propensity Strata: 2008 to 2011 NSDUH 52

5.4 Result Codes for Adults Selected for Clinical Follow-Up Interviews and Agreeing to Participate, 2008 to 2011 53

5.5 Summary of Nonresponse Adjustments for Agreement to Participate in the Clinical Follow-Up, by Propensity Strata: 2008 to 2011 MHSS 57

5.6 Summary of Nonresponse Adjustments for Adults Completing the Clinical Follow-Up Given That They Agreed to be Recontacted, by Propensity Strata: 2008 to 2011 MHSS 61

5.7 Poststratification Controls and Average Adjustment: 2008 to 2011 MHSS Clinical Sample 65

5.8 Unequal Weighting Effects for Weights at Stages of Adjustment, by Year of Data Collection: 2008 to 2012 Adult NSDUH Main Study and Clinical Sample 69

5.9 Direct Estimates Based on SCID Sample: 2008-2012 MHSS 70

6.1 Sample Description of SCID Respondents, by Demographic Characteristics: 2008-2012 81

6.2 Sample Description of SCID Respondents, by Mental Health Characteristics: 2008-2012 82

6.3 Sample Description of SCID Respondents, by Substance Use Measures: 2008-2012 83

6.4 Weighted Sample Description of SCID and CAI Respondents, by Demographic Characteristics: 2008-2012 84

6.5 Weighted Sample Description of SCID and CAI Respondents, by Mental Health Characteristics: 2008-2012 85

6.6 Weighted Sample Description of SCID and CAI Respondents, by Substance Use Measures: 2008-2012 86

List of Tables (continued)

Table **Page**

6.7 Sample Description of SCID Respondents, by Past Year K6 Score Frequency Distribution: 2008-2012 87

6.8 Weighted Sample Description of SCID and CAI Respondents, by Past Year K6 Score Frequency Distribution: 2008-2012 88

6.9 Weighted Sample Description of SCID and CAI Respondents, by K6 Scores: 2008-2012 89

6.10 Weighted WHODAS Past Year Total Score Frequency Distribution of SCID and CAI Respondents, by Sampling Period: 2008-2012 90

7.1 Cut Point Estimates of SMI for Age Groups and Their Bias Measures Using a Model of SMI with Past Year K6, WHODAS, MDE, and Suicidal Thoughts as Predictor Variables: 2008A-2012 MHSS 97

7.2 SMI Cut Point Estimates for All Adults under Different Models: 2008 to 2012 MHSS 98

7.3 Comparing Cut Point Estimates for SMI from the 2008B Sample Using the 2012 Model for the SDS Sample with Estimates from the 2008A Sample Using the 2012 Model for the WHODAS Sample 101

7.4 Comparing Cut Point Estimates for AMI from 2008B Sample Using the 2012 Model for the SDS Sample with Estimates from the 2008A Sample Using the 2012 Model for the WHODAS Sample 102

7.5 Comparing SMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample 104

7.6 Comparing AMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample 106

D1a Descriptive Characteristics, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH D-3

D1b Mental Health Measures, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH D-5

D1c Substance Use Measures, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH D-6

D2a Descriptive Characteristics, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS D-8

D2b Mental Health Measures, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS D-10

D2c Substance Use Measures, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS D-11

List of Tables (continued)

Table **Page**

D3a Descriptive Characteristics, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS D-13

D3b Mental Health Measures, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS D-15

D3c Substance Use Measures, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS D-16

E1 Descriptive Characteristics, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample.....E-3

E2 Mental Health, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample E-6

E3 Substance Use Measures, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample.....E-8

List of Figures

Figure **Page**

3.1 Structure of the 2008 Mental Health Surveillance Clinical Follow-Up Study Sampling Design 14

List of Exhibits

Exhibit **Page**

5.1 Recruitment Screen for 2008-2012 MHSS Clinical Follow-Up 52

1. Introduction

The overarching goal of the Mental Health Surveillance Study (MHSS) of the National Survey on Drug Use and Health (NSDUH) is to provide accurate estimates of the prevalence of serious mental illness (SMI) among adults aged 18 or older at the national and State levels. Public Law No. 102-321, the Alcohol, Drug Abuse, and Mental Health Administration Reorganization Act of 1992, established a block grant for U.S. States to fund community mental health services for adults with SMI. The law required States to include prevalence estimates in their annual applications for block grant funds. This legislation also required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop an operational definition of SMI and to produce national and State estimates. The MHSS clinical follow-up study was conducted to develop a model to generate estimates of SMI. However, the clinical data have the potential to be used for a variety of other purposes beyond this primary purpose, such as estimating other categories of mental illness (e.g., "mild," "moderate," or "any" mental illness).

On May 20, 1993, SAMHSA's Center for Mental Health Services (CMHS) published its definition of SMI in the *Federal Register*:

> Pursuant to Section 1912(c) of the Public Health Services Act, as amended by Public Law 102-321, "adults with serious mental illness" are defined as the following:
>
> - Persons aged 18 and over, who currently or at any time during the past year, have had diagnosable mental, behavioral, or emotional disorder of sufficient duration to meet diagnostic criteria specified within DSM-III-R [sic] that has resulted in functional impairment, which substantially interferes with or limits one or more major life activities.
> - These disorders include any mental disorders (including those of biological etiology) listed in DSM-III-R or their ICD-9-CM equivalent (and subsequent revisions), with the exception of DSM-III-R "V" codes, substance use disorders, and developmental disorders, which are excluded unless they co-occur with other diagnosable serious mental illness.
> - All of these disorders have episodic, recurrent, or persistent features; however, they vary in terms of severity or disabling effects. Functional impairment is defined as difficulties that substantially interfere with or limit role functioning in one or more major life activities, including basic daily living skills (e.g., eating, bathing, dressing); instrumental living skills (e.g., maintaining a household, managing money, getting around the community, taking prescribed medication); and functioning in social, family, and vocational/educational contexts.
> - Adults who would have met functional impairment criteria during the referenced year without benefit of treatment or other support services are considered to have serious mental illnesses.

In December 2006, a technical advisory group meeting of expert consultants was convened by the Office of Applied Studies (OAS, now the Center for Behavioral Health Statistics and Quality [CBHSQ]) and CMHS to solicit recommendations for mental health surveillance data collection strategies among the U.S. population. The panel recommended that

NSDUH be used to produce estimates of SMI among adults by including short scales in NSDUH's main interview that are strong predictors of SMI and that a "gold-standard" clinical psychiatric interview be administered on a subset of respondents to provide the data for estimating a statistical model that predicts SMI. In response, SAMHSA's CBHSQ initiated the MHSS as part of NSDUH to develop and implement a method to estimate SMI. At the time, NSDUH contained a six-item scale (Kessler-6 or K6) with five response options in each item that captured information on psychological distress in the past 12 months (Kessler et al., 2003). However, the K6 scale is not a diagnostic instrument and does not capture information on functional impairment, which is needed to determine whether a respondent can be categorized as having SMI under SAMHSA's definition. In consultation with the technical advisory group, two candidate impairment scales were selected by SAMHSA to be added to the 2008 NSDUH. They were an abbreviated version of the World Health Organization Disability Assessment Schedule (WHODAS; Rehm et al., 1999) and the Sheehan Disability Scale (SDS; Leon, Olfson, Portera, Farber, & Sheehan, 1997). An initial step was to modify these scales for use in a general population survey, including changes to question wording and length, which resulted in an abbreviated eight-item version of the WHODAS (Novak, Colpe, Barker, & Gfroerer, 2010). Further details of the K6 scale are given in Appendix A, and details of the two impairment scales are given in Appendices B and C.

The MHSS clinical interviews were conducted first in 2008. A split-sample design was used in the 2008 NSDUH, for which all adult respondents to the main NSDUH interview received the K6, but a random half of the sample received the WHODAS and the other half received the SDS. In addition, a subsample of approximately 1,500 adult NSDUH participants completed a follow-up clinical interview to provide data for developing models to estimate mental illness using the adult NSDUH data from the main interview. The randomization of the impairment scales was maintained within this clinical interview subsample, which is referred to in this report as the clinical sample, so that about half of the MHSS clinical sample participants were administered the WHODAS and half were administered the SDS (i.e., there were approximately 750 completed interviews from each half sample). Each participant in the 2008 MHSS clinical study was administered the Structured Clinical Interview for DSM-IV (SCI67D),[1] which was adapted for this study by mental health clinicians for paper-and-pencil interviewing administered over the telephone. The clinical interviews were administered approximately 2 to 4 weeks after the main NSDUH interview. Functional impairment ratings were assigned by clinical interviewers using the Global Assessment of Functioning (GAF) scale.[2] The modified SCID questionnaire for the 2008-2012 MHSS clinical study is shown in Appendix C of the 2008-2012 MHSS operations report (CBHSQ, 2014a). The model estimation analyses used gold-standard measures (i.e., the SCID/GAF combination as the indicator of SMI) in evaluating which

[1] The Structured Clinical Interview for the DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP) (First, Spitzer, Gibbon, & Williams, 2002).

[2] The GAF is a numeric scale (0 through 100) used to subjectively rate the social, occupational, and psychological functioning of adults and is presented and described in the DSM-IV-TR (see p. 32 of American Psychiatric Publishing, Inc., 2000; also see Endicott, Spitzer, Fleiss, & Cohen, 1976). Lower scores represent higher levels of functional impairment. Descriptions of impairment are provided at 10-point intervals (e.g., 1 to 10, 11 to 20, and so on up to 91 to 100). For example, a GAF score between 51 and 60 is described as having moderate symptoms of impairment, while a score higher than 60 represents several categories of impairment ranging from none to slight, and a score lower than 51 represents several categories ranging from serious to extreme.

combination of K6 and impairment scale worked best in the statistical model used to predict SMI status.

Based on an analysis of the 2008 MHSS data, it was determined that the WHODAS would be administered as the sole impairment scale in subsequent NSDUHs (starting in 2009) and that it would be used in combination with the K6 scale to predict SMI. For more details, refer to the 2008 MHSS analysis report by Aldworth et al. (2009). From 2009 through 2012, the MHSS was conducted similarly to the 2008 MHSS, except for two major differences: (1) only the WHODAS impairment scale was administered, and (2) the sample size was approximately 500 in 2009 and 2010, and the sample size was approximately 1,500 in 2011 and 2012.

The primary objective of the MHSS analysis is to produce annual national estimates of SMI prevalence that are accurate for all adults and for adult subpopulations. Secondary objectives include producing estimates of other categories of mental illness defined by level of impairment, such as mild (or low) mental illness (LMI), moderate mental illness (MMI), and any mental illness (AMI). These categories of mental illness, which are based on SCID disorder diagnoses and GAF scores, are defined in Table 1.1. A respondent was coded positive for SMI if he or she was determined to have any of the mental disorders (not including developmental or substance use disorders) assessed in the MHSS SCID *and* had a GAF score of 50 or below. AMI, defined as having a mental disorder regardless of the level of impairment due to that disorder, is the category obtained by collapsing the first three categories in Table 1.1 into a single category.

Table 1.1 Mental Illness Categories Defined by SCID Disorder Diagnosis and GAF Score: 2008-2012 MHSS Clinical Follow-Up Study

Mental Illness Category	SCID Disorder Diagnosis	GAF Score
Serious Mental Illness (SMI)	One or More	GAF ≤ 50 (severe or worse impairment)
Moderate Mental Illness (MMI)	One or More	50 < GAF ≤ 59 (moderate impairment)[1]
Mild (or Low) Mental Illness (LMI)	One or More	59 < GAF (at most mild impairment)
No Mental Illness (No MI)	None	GAF score not applicable

GAF = DSM-IV Axis V Global Assessment of Functional Scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition.

[1] DSM-IV description of moderate impairment based on GAF is 50 < GAF ≤ 60. The cutoff of 59 for MMI and LMI was chosen to conform to the corresponding cutoff selected by Kessler et al. (2003).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH MHSS clinical sample, 2008-2012.

A model was developed based on the 2008 clinical data and used in the 2008 through 2011 NSDUHs to produce a predicted probability of having SMI for each clinical interview respondent. A cut point was established among the fitted probabilities of having SMI based on the 2008 MHSS clinical data, such that adults with probabilities at or above the cut point were predicted to have SMI and the rest were not. Although this model was optimized to predict SMI, the SMI predicted probabilities were also used to predict MMI and LMI using different cut points (for more details about this model, see Liao et al., 2012).

With the accumulated MHSS clinical data collected from 2008 to 2012, SAMHSA determined that the 2008 model had some important shortcomings that had not been detected in the original model fitting because of the small number of respondents in the 2008 clinical sample. Specifically, the 2008 model substantially overestimated SMI and AMI among young adults in NSDUH relative to young adults in the clinical interview data. In addition, improvements were needed in the weighting procedures for the MHSS clinical data to account better for undercoverage and nonresponse (i.e., because only NSDUH respondents who answered their surveys in English were eligible for the clinical follow-up and because persons with mental illness appeared to be more likely to participate in the follow-up). Therefore, using the combined 2008-2012 clinical data, SAMHSA fit a more accurate model for the 2012 estimates with revised weights (subsequently referred to as the "2012 model" and "2012 estimation methods"). In particular, to reduce bias and improve prediction, additional mental health-related variables and an age variable were added in the 2012 model. In addition, to protect against potential coverage and nonresponse error, alternatives for the weights were applied to the clinical sample data for the model development. To provide consistent data for trend assessment, mental illness estimates for 2008 to 2011 using the new 2012 model were revised.

The remainder of this report is organized into seven chapters. Chapter 2 describes the mental illness and impairment scales and instruments that are used to produce SMI estimates. Chapter 3 discusses the sample design and methods for selecting respondents for the MHSS clinical interview. Chapter 4 illustrates how the collected MHSS data are processed for analytical purposes, including coding, editing, and imputation for the missing data. Chapter 5 clarifies the components of the MHSS analysis weights, including the methodology developed to adjust for nonresponse and noncoverage errors, while simultaneously preventing extreme weights. Chapter 6 summarizes the results of the descriptive analyses that compare the key demographic and psychosocial characteristics across samples from different time periods. Chapter 7 provides a general overview of the estimation methodology used to produce the prevalence estimates of mental illness among adults aged 18 or older and describes the revisions to the estimation methodology that were implemented in 2012.

2. Instruments Used in Measuring Mental Illness

2.1 Background

This chapter describes the psychological distress and impairment scales, as well as the clinical instrument, that were administered to adult respondents and used to produce estimates of serious mental illness (SMI). All adult respondents aged 18 or older in the National Survey on Drug Use and Health (NSDUH) were asked about their level of psychological distress. If a respondent had a psychological distress score greater than zero, he or she was then directed to questions on functional impairment.

In the main NSDUH interview, the Kessler-6 (K6) scale was used to assess psychological distress. In the 2008 NSDUH, a split-sample design was implemented in which all adult respondents in the main interview received the K6, but a random half of the sample received the World Health Organization Disability Assessment Schedule (WHODAS) and the other half received the Sheehan Disability Scale (SDS) for their impairment assessment. From the 2009 survey onward, the WHODAS was administrated as the sole impairment scale for adult respondents. The sample size was approximately 500 in 2009 and 2010 and approximately 1,500 in 2008, 2011, and 2012.

Each survey year from 2008 to 2012, a subsample of the adult NSDUH participants participated in a follow-up clinical interview and were administered the Structured Clinical Interview for DSM-IV (SCID) (First et al., 2002), including a module assessing Axis I disorders, and the Global Assessment of Functioning (GAF) scale by trained clinical interviewers.

2.2 Scales in the Main NSDUH Interview

2.2.1 Psychological Distress Scale (K6)

The K6 scale, used to capture nonspecific psychological distress (Kessler et al., 2003), consists of two sets of six questions that ask respondents how frequently they experienced symptoms of psychological distress during two different time periods: (1) during the past 30 days and (2) the one month in the past year when they were at their worst emotionally. Respondents were only asked about the second time period if they indicated that there was a month in the past 12 months when they felt more depressed, anxious, or emotionally stressed than they felt during the past 30 days. The six domains covered by the questions corresponded to how often the respondent felt (1) nervous, (2) hopeless, (3) restless or fidgety, (4) sad or depressed, (5) that everything was an effort, and (6) worthless. To create a score, the six items related to the first time period were coded from 0 to 4 so that "all of the time" was coded 4, "most of the time" 3, "some of the time" 2, "a little of the time" 1, and "none of the time" 0, with "don't know" and "refuse" also coded as 0. Summing across the six responses resulted in a total score with a range from 0 to 24. The six items related to the second time period were coded identically, and the worst K6 total score was calculated as the maximum of the total scores from the two time periods and is considered the past year K6 total score. An alternative version of the past year K6

total score was formulated as follows: past year K6 total scores of less than 8 were recoded as 0, and past year K6 total scores from 8 to 24 were recoded as 1 to 17. The reason for the alternative version was that the SMI prevalence was typically extremely low for respondents with past year K6 total scores of less than 8, and the prevalence rates were higher, in general, for total scores of 8 or greater. See Appendix A for the specific K6 scale items.

2.2.2 Functional Impairment Scales (WHODAS and SDS)

The abbreviated WHODAS, used to capture impairment data (Rehm et al., 1999), consists of eight questions that ask respondents how much their emotions, nerves, or mental health caused them to have difficulties in daily activities over the past year (Novak et al., 2010). Eight domains were covered by the following questions:

1. remembering to do things they needed to do,
2. concentrating on doing something important when other things were going on around them,
3. going out of the house and getting around on their own,
4. dealing with people they did not know well,
5. participating in social activities,
6. taking care of household responsibilities,
7. taking care of daily responsibilities at work or school, and
8. getting daily work done as quickly as needed.

To create a score, the eight items were coded from 0 to 3 so that "severe difficulty" was coded 3, "moderate difficulty" 2, "mild difficulty" 1, and "no difficulty" 0, with "don't know" and "refuse" also coded as 0. Some items had a fifth category to address "not applicable" responses. For example, the question about difficulties regarding taking care of daily responsibilities at work or school had a fifth category, "you didn't go to work or school." If this category was selected, then a further question was asked as to whether their emotions, nerves, or mental health caused them to be unable to go to work or school. A "yes" response to the follow-up question was coded 3, and a "no" response was coded 0.

One exception to this coding related to the last WHODAS item on how much difficulty respondents had in getting their daily work done as quickly as needed. This item was only asked if in the previous item on assessing their ability to function at work or school they responded to any of the first four categories (i.e., implying that they went to work or school) and was coded similarly to the other items. If they responded to the fifth category (i.e., that they did not go to work or school), their response to this item was determined by the final code for the follow-up item on whether their emotions, nerves, or mental health caused them to be unable to go to work or school.

Summing across the eight responses resulted in a total score with a range from 0 to 24. An alternative version of the WHODAS total score was formulated as follows: item scores of less than 2 were recoded as 0, and item scores from 2 to 3 were recoded as 1, then summed for a

total score ranging from 0 to 8. The reason for creating an alternative version of the WHODAS total score was the notion that using a dichotomous measure dividing respondents who experienced moderate or severe difficulties from the remaining respondents might produce a better predictor of SMI than would a linear continuous measure. See Appendix B for the actual questions used in the WHODAS.

The SDS, used only in half of the 2008 NSDUH sample to capture impairment data (Leon et al., 1997), consists of four questions that ask respondents how much their emotions, nerves, or mental health interfered with their daily activities over the past year. The following four domains were covered by the questions: (1) home management, (2) work, (3) close relationships with others, and (4) social life. For each of the four items, respondents were asked to select a number from 0 to 10 on a visual analog scale, where 0 means no interference, 1 to 3 mild interference, 4 to 6 moderate interference, 7 to 9 severe interference, and 10 very severe interference. Summing across the four responses resulted in a total score with a range from 0 to 40. An alternative version of the SDS total score was formulated as follows: item scores of less than 7 were recoded as 0, and item scores from 7 to 10 were recoded as 1, then summed for a total score ranging from 0 to 4. The alternative version of the SDS total score was again driven by the notion that a dichotomous measure dividing respondents who experienced severe or very severe interference from the remaining respondents might produce a better predictor of SMI than would a linear continuous measure. See Appendix C for the actual questions used in the SDS.

2.3 Clinical Follow-Up Instruments

Each participant in the 2008 to 2012 MHSS clinical follow-up study was administered standard clinical interview measures by mental health clinicians via paper-and-pencil interviewing over the telephone within 2 to 4 weeks of the NSDUH main interview. The MHSS clinical interview measure was the SCID (First et al., 2002), which is a semistructured diagnostic interview used to assess psychiatric disorders according to the criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994). As a semistructured clinical interview, the SCID contains structured, standardized questions that are read verbatim and sequentially, combined with unstructured follow-up questions that the clinical interviewer tailors to the respondent based on clinical judgment and respondent reports. The SCID was modified for the MHSS to assess specific mental disorders and functioning in the past 12 months. The SCID was administered over the telephone by trained clinical interviewers. The SCID has been widely used in clinical calibration studies, such as the National Comorbidity Survey Replication (NCS-R; Kessler et al., 2004), the National Survey of American Life (Jackson, Neighbors, Nesse, Trierweller, & Torres, 2004), and NSDUH's substance use disorders reappraisal study (Jordan, Karg, Batts, Epstein, & Wiesen, 2008). It has demonstrated good reliability (Segal, Kabacoff, Hersen, Van Hasselt, & Ryan, 1995; Zanarini et al., 2000; Zanarini & Frankenburg, 2001) and validity (Fennig, Craig, Lavelle, Kovasznay, & Bromet, 1994; Kranzler, Kadden, Babor, Tennen, & Rounsaville, 1996; Kranzler et al., 1995; Ramirez Basco et al., 2000; Shear et al., 2000; Steiner, Tebes, Sledge, & Walker, 1995). Studies that compared telephone versus face-to-face administration of the SCID have found good agreement (Crippa et al., 2008; Hajebi et al., 2012; Kendler, Neale, Kessler, Heath, & Eaves, 1992; Kessler et al., 2004; Lee et al., 2008; Rohde, Lewinsohn, & Seely, 1997; Sobin et al., 1993).

Diagnostic modules in the MHSS version of the SCID are listed in Table 2.1. The assessment of lifetime manic episode was included to provide a context for understanding whether a past 12-month major depressive episode (MDE) was experienced as part of a unipolar mood disorder or as a component of a bipolar disorder (regardless of whether a manic episode also was experienced in the past year). The module to assess intermittent explosive disorder was obtained from the (optional) impulse control disorders section of the SCID. Although the module for substance use disorders was administered to respondents, substance use disorder is not included in SAMHSA's definition of SMI and was therefore not used in the estimation of SMI.

Table 2.1 Diagnostic Modules in the 2008-2012 MHSS SCID

MOOD DISORDERS	PAST YEAR EATING DISORDERS
Past Year Major Depressive Episode[1]	Anorexia Nervosa[1]
Lifetime Major Depressive Episode	Bulimia Nervosa[1]
Past Year Manic Episode[1]	
Lifetime Manic Episode	PAST YEAR IMPULSE CONTROL DISORDERS
Dysthymic Disorder[1]	Intermittent Explosive Disorder[1]
PAST YEAR PSYCHOTIC DISORDERS	
Psychotic Screen[1]	PAST YEAR SUBSTANCE USE DISORDERS
	Alcohol Abuse
PAST YEAR ANXIETY DISORDERS	Alcohol Dependence
Posttraumatic Stress Disorder[1]	Non-Alcohol Substance Abuse
Panic Disorder with and without Agoraphobia[1]	Non-Alcohol Substance Dependence
Agoraphobia without History of Panic Disorder[1]	
Social Phobia[1]	PAST YEAR ADJUSTMENT DISORDERS
Specific Phobia[1]	Adjustment Disorder[1]
Obsessive Compulsive Disorder[1]	
Generalized Anxiety Disorder[1]	

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition.

[1] Disorder used to determine gold-standard measures of serious mental illness and other categories of mental illness.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH MHSS clinical sample, 2008-2012.

In addition to the diagnostic modules, the MHSS SCID included four other sections:

1. An open-ended overview module is part of the standard SCID designed to elicit information about the respondent's diagnostic and treatment history and current status in a way that establishes some level of rapport between the clinical interviewer and the respondent.
2. The SCID screener instrument is a set of questions at the end of the overview section and is administered to all respondents. Its questions are taken from the body of the SCID and are the initial questions asked by the SCID for the disorders being assessed. These screening questions may help to reduce the potential effects of a "negative response bias" that may be especially problematic in the later sections of the SCID. Because of the structure of the SCID, there is a tendency for the respondent to notice that a "yes" answer to the initial probe question in a section results in follow-up

questions, whereas a "no" answer results in a skip to the next section, thus leading some subjects to give "no" answers as a way of speeding the interview along. By asking these questions up front and using the answers to these questions in the determination of whether a section should be skipped, response bias may be minimized.

3. With the DSM-IV Axis V GAF scale, the clinical interviewer was instructed to rate the respondent's period of worst psychological, social, and occupational functioning during the past year.
4. A section is included for documenting the clinical interviewer's impressions of the interview situation, including ratings of the respondent's level of privacy, cooperation, and comprehension, as well as the overall validity of the interview data (any interview deemed by the clinical interviewer or clinical supervision team to be of questionable validity was discarded).

For more details about these modules, see Colpe et al. (2010). Also, see Appendix C of the 2008-2012 MHSS operations report for the full modified SCID questionnaire (Center for Behavioral Health Statistics and Quality, 2014a).

3. Sample Design and Selection

3.1 Background

This chapter describes the sample design and methods for selecting respondents for the Mental Health Surveillance Study (MHSS) clinical interview. The clinical sample was selected from the main National Survey on Drug Use and Health (NSDUH) study sample of approximately 45,000 adults. The target population included those who completed the NSDUH main interview in English and excluded persons who completed the NSDUH main interview in Spanish. A probability sampling algorithm was programmed in the computer-assisted interviewing (CAI) instrument such that field interviewers (FIs) could, at the conclusion of the interview, recruit selected respondents for the subsequent clinical psychiatric interview conducted by telephone. This chapter describes, in detail, the annual MHSS clinical sample selection process, which began with a feasibility study in 2007 and evolved over the 2008 to 2012 data collection periods. Other issues related to the sample selection also are described. Descriptions of the 2008 to 2012 clinical samples are provided in Chapter 6.

3.2 Selecting the MHSS Clinical Follow-Up Samples

NSDUH respondents (i.e., NSDUH main interview respondents) aged 18 or older who completed their interviews in English were eligible to be sampled for the MHSS clinical follow-up study. The procedures used to identify which NSDUH respondents would be selected for the clinical interview varied over the data collection period. Prior to the initiation of the MHSS clinical follow-up study, a feasibility study was conducted in 2007 in which respondents were selected to complete the NSDUH main interview first, and all respondents who completed the main interview in English were selected for the clinical follow-up. Details about the 2007 feasibility study are discussed in Section 3.3.

In 2008 and 2009, the probability of selection for the clinical follow-up study was based on the respondent's Kessler-6 (K6) score. The K6 score was used because, as demonstrated by data from the National Comorbidity Survey Replication (NCS-R) clinical calibration study,[3] it is highly correlated with serious mental illness (SMI), the key characteristic of interest in the MHSS. Based on observed variation in the weights in the 2008 and 2009 MHSS clinical studies that affected the precision of resulting estimates, a design change was implemented starting in 2010. In 2010 through 2012, in order to decrease the variability of the weights, the probability of selection was based on the respondent's age, K6 score, and World Health Organization Disability Assessment Schedule (WHODAS) score. Because the target population and sampling frames for the 2008 through 2012 MHSS clinical follow-up studies remained the same, these changes in the allocation of selected persons did not affect the point estimates themselves. Instead, these design changes allowed for more precise estimates because they resulted in less variable analysis weights. Details of the selection probabilities for the MHSS clinical sample are discussed in Sections 3.4, 3.5, and 3.6.

[3] Kessler, R. C., attachment to a personal email communication to L. J. Colpe, August 1, 2007, Scidsmi-table-073107 (2) (2).doc.

3.3 2007 Feasibility Study

A feasibility study was conducted in June 2007 to test the procedures and instruments proposed for the 2008 MHSS clinical follow-up study. The feasibility study was conducted in order to assess all of the procedures necessary for conducting and reporting on the clinical interviews, as well as to assess the training needs for both the FIs and the clinical interviewers.

Unlike the 2008 through 2012 clinical studies, respondents to the clinical follow-up interview for the feasibility study were not selected from the main NSDUH study sample of adults. Instead, a new sample of dwelling units was selected from 10 sample segments of the 2006 NSDUH that had been retired from use in the main study. These retired segments were included in the 2005 and 2006 NSDUH samples and would not be reused in future NSDUH survey years. The segments were selected based on their proximity to experienced NSDUH FIs who were available to work on the feasibility study.

Consistent with the main NSDUH study, sampled dwelling units in the feasibility study were fielded and screened for main study eligibility. Within sampled dwelling units, zero, one, or two adults were selected for the initial NSDUH in-person interview. All respondents selected for and completing the initial interview in English were selected for the clinical follow-up telephone interview.

In preparation for the feasibility study, it was estimated that 270 dwelling units would be needed to yield 48 completed clinical interviews (see Table 3.1). Based on national-level experience from the 2006 NSDUH, adjusted for the shorter than normal data collection period,

Table 3.1 2007 Mental Health Surveillance Feasibility Study Design Parameters

Design Parameters	Count/Percent
Selected for the Initial Main Interview	
Selected Segments	10
Dwelling Units per Segment	27
Total Selected Dwelling Units	270
Dwelling Unit Eligibility Rate	83%
Eligible Dwelling Units	224
Expected Screening Response Rate	90%
Completed Screenings	202
Expected Yield 18 or Older	40%
Sampled Persons Aged 18 or Older	81
Assumed Interview Response Rate	70%
Selected for Telephone Clinical Follow-Up, by SPD Score	
Score 0 to 7	37, (66%)
Score 8 to 12	11, (19%)
Score 13 or Higher	8, (15%)
Total Selected for Telephone Clinical Follow-Up	56
Clinical Follow-Up Interview Response Rate	85%
Completed Clinical Follow-Up Interviews	48

NSDUH = National Survey on Drug Use and Health; SPD = serious psychological distress.

NOTE: Estimated sample proportions, eligibility rate, and response rates were derived from 2006 NSDUH data, adjusted for the shorter than normal data collection period.

an 83 percent dwelling unit eligibility rate, a 90 percent screening response rate, a 70 percent interview response rate, and an 85 percent clinical follow-up interview response rate were assumed. Also based on 2006 NSDUH data, it was expected that 66 percent of the interview respondents would have a serious psychological distress (SPD) score between 0 and 7, 19 percent would have a score between 8 and 12, and 15 percent would have a score of 13 or higher.

In the sample selection implementation, a total of 262 dwelling units were selected from the 10 sample segments. The dwelling unit sample size and person-level probabilities of selection were determined using sample allocation procedures identical to those for the 2007 NSDUH main study. State-specific eligibility rates, response rates, and yields were estimated using data from the 2006 NSDUH and adjusting for the shorter than normal data collection period. Using this State-specific information, fewer selected dwelling units were required than was originally anticipated. To ensure that the targeted number of clinical interviews was achieved with the feasibility study sample, a 20 percent reserve sample was also selected. To allow for the 20 percent reserve sample to be released as needed, the total sample of 262 dwelling units was partitioned into two releases within each segment.

A total of 232 sample dwelling units were fielded, 216 of which were eligible, giving a dwelling unit eligibility rate of 93.1 percent. Household screenings were completed at 195 households, for an unweighted screening response rate of 90.3 percent. A total of 102 sample persons were selected within these responding dwelling units, and interviews were completed with 70 for an unweighted initial interview response rate of 68.6 percent. Out of the 70 respondents who completed the initial main NSDUH field interview, 60 agreed to complete the follow-up clinical telephone interview. The unweighted percentage of respondents *agreeing* to complete the feasibility study clinical interview was 85.7 percent. Out of the 60 respondents who agreed to complete the clinical interview, 45 clinical interviews were completed. The unweighted percentage of respondents completing clinical interviews was 75.0 percent. The results of the feasibility study were used to inform the procedures utilized in the 2008 MHSS clinical study.

3.4 2008 Sample Allocation

The 2008 clinical data collection included 1,500 clinical follow-up interviews. In the 2008 NSDUH main study, adult respondents were randomly assigned to one of two functional impairment scales: an abbreviated version of the WHODAS (Novak et al., 2010; Rehm et al., 1999) and the Sheehan Disability Scale (SDS) (Leon et al., 1997). Approximately equal numbers of respondents received each of the functional impairment scales, and all of the respondents received the K6 items. The randomization of the impairment scales was maintained within the MHSS clinical sample so that about half of the MHSS clinical interview respondents (approximately 750) were administered the WHODAS and half were administered the SDS. A diagram illustrating the structure of the 2008 MHSS clinical study sampling design is given in Figure 3.1.

Figure 3.1 Structure of the 2008 Mental Health Surveillance Clinical Follow-Up Study Sampling Design

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008.

NSDUH respondents eligible for the clinical follow-up were selected for the clinical interview through stratified random sampling. To achieve an efficient sample design, sampling strata were formed by grouping K6 scores with similar rates of SMI based on data collected from the NCS-R clinical calibration study. Several stratification options were considered in the optimization process, and a stratified sample design with seven sampling strata based on their K6 scores (0 to 3, 4 to 5, 6 to 7, 8 to 9, 10 to 11, 12 to 15, and 16 or higher) was adopted. This option was selected because it minimized the projected relative standard error (RSE) and more effectively smoothed out the sample distribution across K6 scores 4 through 11 compared with allocations based on fewer strata.

To optimize the MHSS clinical sample allocation for the clinical follow-up within scoring bands, assumed SMI rates were estimated using K6 scores and clinical case data from the NCS-R clinical calibration study.[4] Assumed SMI rates for the 2008 study were set equal to the NCS-R rates except in the instance where the K6 scores ranged from 0 to 7. In this instance, SMI rates were set lower under the assumption that fewer clinical positives would be identified in that scoring range. Population percentages by K6 group were estimated from the 2006 NSDUH.

[4] The design of the 2008 MHSS clinical study was based on data from the NCS-R clinical calibration study because these were the best estimates available at that time. Later designs based their assumptions on results from the prior year's MHSS clinical follow-up study.

Using Neyman's optimal allocation (Lohr, 1999), a solution that minimized the design effect for the prevalence of SMI in the clinical sample was selected. Table 3.2 shows the expected sample distribution for the 1,500 clinical follow-up interviews. The expected overall design effect[5] was 0.6363, which led to an expected effective sample size of 2,357. The projected standard error (SE) and RSE of the all-adult estimate of SMI prevalence based on the MHSS clinical sample were 0.59 and 6.57 percent, respectively.

Table 3.2 2008 Mental Health Surveillance Clinical Follow-Up Study Sample Allocation (*n* = 1,500)

K6 Score	Percent of Population	Assumed SMI Rate (%)	Sample Size
0 to 3	48.04	0.03	96
4 to 5	13.98	0.30	88
6 to 7	11.16	0.30	110
8 to 9	6.95	10.00	200
10 to 11	5.53	13.00	214
12 to 15	8.00	40.00	450
16 or Higher	6.34	67.00	343
Total	**100.00**	**8.95**	**1,501**[1]

K6 = Kessler-6, a 6-item psychological distress scale; NCS-R = National Comorbidity Survey Replication; NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness.

[1] This number does not equal the number for the sample size due to rounding.

NOTE: The population source is the 2006 NSDUH. Assumed SMI rates were estimated using data from the 2001-2002 NCS-R clinical calibration study.

The NSDUH CAI instrument included a sampling algorithm to indicate whether NSDUH main interview respondents had been sampled for the clinical follow-up study. FIs recruited sampled respondents for the subsequent clinical psychiatric interview that was conducted by telephone.

The probability sample of 1,500 clinical follow-up interviews was distributed across four calendar quarters with a slightly larger sample in the first quarter (425 follow-up interviews) and the remaining sample equally divided among the remaining quarters (approximately 358 interviews in each of the quarters 2 through 4 for a combined sample of 1,075 clinical follow-up interviews; see Table 3.3). The larger sample in quarter 1 was intended to provide some cushion should the clinical interview response rates be lower than anticipated. In addition, a slightly larger sample size in quarter 1 was needed to allow for preliminary analyses of the data. The sample sizes were determined based on an assumed 96 percent clinical interview eligibility rate,[6] an 85 percent agreement rate for the clinical follow-up interview, and a 90 percent participation rate among those who agreed to complete the interview.

[5] The expected design effect for the 2008 MHSS clinical follow-up study is the product of the usual design effect for adults in the main survey (about 3.0) and the design effect for the two-phase sample stratified by K6 scores (about 0.2).

[6] The MHSS clinical interview eligibility rate was not accounted for in the quarter 1 2008 design, but it was incorporated in the sample designs for quarters 2 to 4 in 2008.

Table 3.3 Design Parameters for the 2008 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4

Design Parameter	Quarter 1 Total	Quarters 2 through 4 Total	Quarter 1 Expected Persons per Segment	Quarters 2 through 4 Expected Persons per Segment
Interview Respondents Aged 18 or Older, by K6 Score	11,250	33,750	6.3	6.3
Score 0 to 3 (42% of Cases)	4,725	14,175	2.6	2.6
Score 4 to 5 (13% of Cases)	1,463	4,388	0.8	0.8
Score 6 to 7 (12% of Cases)	1,350	4,050	0.8	0.8
Score 8 to 9 (8% of Cases)	900	2,700	0.5	0.5
Score 10 to 11 (6% of Cases)	675	2,025	0.4	0.4
Score 12 to 15 (10% of Cases)	1,125	3,375	0.6	0.6
Score 16 or Higher (9% of Cases)	1,013	3,038	0.6	0.6
Sampling Rate, by K6 Score				
Score 0 to 3	0.0075	0.0063		
Score 4 to 5	0.0223	0.0188		
Score 6 to 7	0.0301	0.0254		
Score 8 to 9	0.0823	0.0694		
Score 10 to 11	0.1174	0.0990		
Score 12 to 15	0.1481	0.1249		
Score 16 or Higher	0.1255	0.1058		
Selected for Telephone Clinical Follow-Up, by K6 Score				
Score 0 to 3	35	89	0.0	0.0
Score 4 to 5	33	82	0.0	0.0
Score 6 to 7	41	103	0.0	0.0
Score 8 to 9	74	187	0.0	0.0
Score 10 to 11	79	200	0.0	0.0
Score 12 to 15	167	422	0.1	0.1
Score 16 or Higher	127	321	0.1	0.1
Total Selected for Telephone Clinical Follow-Up	556	1,405	0.3	0.3
Percent Agreeing to Clinical Follow-Up	0.85	0.85		
Percent Completing the Clinical Follow-Up Interview	0.9	0.9		
Completed Clinical Follow-Up Interview, by K6 Score	425	1,075	0.2	0.2
Score 0 to 3	27	68	0.0	0.0
Score 4 to 5	25	63	0.0	0.0
Score 6 to 7	31	79	0.0	0.0
Score 8 to 9	57	143	0.0	0.0
Score 10 to 11	61	153	0.0	0.0
Score 12 to 15	127	322	0.1	0.1
Score 16 or Higher	97	246	0.1	0.0

K6 = Kessler-6, a 6-item psychological distress scale; NSDUH = National Survey on Drug Use and Health.

NOTE: The estimated distribution of interview respondents is based on the 2006 NSDUH.

Throughout the 2008 survey, the MHSS clinical sample was monitored, and the sampling parameters were modified on an as-needed basis to ensure that the goal of 1,500 completions was achieved. In addition, for the last 4 weeks in quarter 4,[7] the probability of selection of the NSDUH respondents for the clinical follow-up survey was set to zero so that cases would not be sampled without adequate time for completing the collection of the interviews (by December 19, 2008) (see Section 3.8.1 for further details).

[7] The MHSS clinical interview recruitment in 2008 ended on November 21st, the MHSS clinical study data collection ended on December 19th, and the NSDUH main study data collection ended on December 22nd.

At the end of the 2008 data collection period, 86 percent of the persons selected from NSDUH agreed to participate in the MHSS clinical follow-up, and 76 percent of those persons completed the MHSS clinical interview. The 2008 MHSS clinical study resulted in 1,500 completed clinical interviews, excluding cases removed from the dataset due to data errors (see Section 3.8.4) or unusable cases (see Section 3.8.5). A summary of the 2008 MHSS clinical interview respondents by quarter is included in Table 3.4.

Table 3.4 2008 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,692	12,816	11,355	10,815	45,678
Unweighted K6 Distribution, by K6 Score					
Score 0 to 3	0.45	0.45	0.45	0.44	0.45
Score 4 to 5	0.15	0.14	0.14	0.15	0.14
Score 6 to 7	0.10	0.10	0.10	0.10	0.10
Score 8 to 9	0.07	0.07	0.07	0.08	0.07
Score 10 to 11	0.05	0.06	0.06	0.06	0.06
Score 12 to 15	0.09	0.09	0.09	0.09	0.09
Score 16 or Higher	0.09	0.09	0.09	0.09	0.09
Eligible for MHSS	10,215	12,148	10,849	10,381	43,593
Eligibility Rate	*0.9554*	*0.9479*	*0.9554*	*0.9599*	*0.9544*
Selected for Telephone Clinical Follow-Up[1]	696	529	485	621	2,331
Zero Probability Cases	0	0	0	47	47
Agreed to Clinical Follow-Up	586	462	416	509	1,973
Percent Agreeing to Clinical Follow-Up (Including Zero Probability Cases)	*0.8420*	*0.8733*	*0.8577*	*0.8196*	*0.8464*
Percent Agreeing to Clinical Follow-Up (Excluding Zero Probability Cases)	*0.8420*	*0.8733*	*0.8577*	*0.8868*	*0.8638*
Completed Clinical Follow-Up Interviews	467	361	317	355	1,500
Clinical Follow-Up Interview Completion Rate	*0.7969*	*0.7814*	*0.7620*	*0.6974*	*0.7603*

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

[1] Includes cases assigned a zero probability of selection that would have been selected based on their K6 rates.

NOTE: The overall response rate to the clinical follow-up study should also consider the response rate to the main NSDUH. In 2008, the weighted overall response rate for the main study was 65.79 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008.

3.5 2009 Sample Allocation

The 2009 data collection was designed to yield 500 clinical follow-up interviews. A subsample of eligible respondents was selected for clinical follow-up with probabilities based on their K6 scores. Similar to the 2008 CAI instrument, the CAI instrument in 2009 included a sampling algorithm to indicate to an FI whether a NSDUH main interview respondent had also been selected for the clinical follow-up survey.

The 2009 sample was initially allocated to seven K6 scoring bands in the same proportions as the 2008 sample. Midway through the year, the decision was made to allocate the sample based on any mental illness (AMI) rather than SMI in order to reduce the selection

probabilities of respondents with extremely large weights while maintaining the efficiency of estimating SMI and AMI directly based on the clinical sample. A new allocation scheme was developed through Neyman's optimal allocation (Lohr, 1999) using estimated population percentages from the 2008 NSDUH main study and AMI estimates from the 2008 MHSS clinical study. This allocation minimized the projected SE and RSE of AMI estimates. Because AMI is prevalent in every K6 scoring band, the new allocation increased the sampling rates in the lower K6 ranges and therefore reduced the size of the weights in those K6 groups. Table 3.5 shows the expected sample distribution for the 500 clinical follow-up interviews under this modified design (i.e., based on SMI in quarters 1 and 2 and based on AMI in quarters 3 and 4). The expected overall design effect[8] was 2.0072, which led to an expected effective sample size of 249. The projected SE and RSE of the all-adult estimate of SMI prevalence based on the MHSS clinical sample were 2.39 and 13.92 percent, respectively.

Table 3.5 2009 Mental Health Surveillance Clinical Follow-Up Study Sample Allocation: Modified Design (*n* = 500)

K6 Score	Percent of Population	Assumed AMI Rate (%)	Sample Size
0 to 3	53.10	3.00	96
4 to 5	13.98	13.42	57
6 to 7	9.35	13.95	47
8 to 9	6.08	33.84	59
10 to 11	4.52	43.43	59
12 to 15	6.77	53.78	103
16 or Higher	6.21	76.04	79
Total	**100.00**	**17.15**	**500**

AMI = any mental illness; K6 = Kessler-6, a 6-item psychological distress scale; NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008.

The probability sample of 500 clinical follow-up interviews was distributed across four calendar quarters with approximately 125 follow-up interviews per quarter. Based on response rate estimates over quarters 1 through 4 of the 2008 MHSS, a 96 percent clinical interview eligibility rate, an 86 percent agreement rate for the clinical follow-up interview, and a 76 percent participation rate among those who agreed to complete the interview were assumed in the sample selection determinations. Table 3.6 displays the 2009 design parameters for quarters 1 through 4. Throughout 2009, the sample was monitored, and the sampling parameters were modified quarterly to ensure that the goal of 500 completions was achieved. In addition, for the last 5 weeks in quarter 4,[9] the probability of selection of the NSDUH respondents for the clinical follow-up survey was set to zero so that cases would not be sampled without adequate time for completion (by December 14, 2009) (see Section 3.8.1 for further details).

[8] The expected design effect for the 2009 MHSS clinical study was the product of the usual design effect for adults in the main survey (about 3.0) and the design effect for the two-phase sample stratified by K6 scores (about 0.7).

[9] The MHSS clinical interview recruitment in 2009 ended on November 16th, the MHSS clinical study data collection ended on December 14th, and the NSDUH main study data collection ended on December 21st.

Table 3.6 Design Parameters for the 2009 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4

Design Parameter	Quarters 1 and 2 Total	Quarters 3 and 4 Total	Quarters 1 and 2 Expected Persons per Segment	Quarters 3 and 4 Expected Persons per Segment
Interview Respondents Aged 18 or Older, by K6 Score	22,500	22,500	6.250	6.250
Score 0 to 3 (45% of Cases)	10,150	10,150	2.820	2.820
Score 4 to 5 (14% of Cases)	3,205	3,205	0.890	0.890
Score 6 to 7 (10% of Cases)	2,298	2,298	0.638	0.638
Score 8 to 9 (7% of Cases)	1,573	1,573	0.437	0.437
Score 10 to 11 (6% of Cases)	1,248	1,248	0.347	0.347
Score 12 to 15 (9% of Cases)	1,977	1,977	0.549	0.549
Score 16 or Higher (9% of Cases)	1,978	1,978	0.549	0.549
Sampling Rate, by K6 Score[1]				
Score 0 to 3	0.0024	0.0076		
Score 4 to 5	0.0074	0.0138		
Score 6 to 7	0.0130	0.0167		
Score 8 to 9	0.0336	0.0300		
Score 10 to 11	0.0503	0.0379		
Score 12 to 15	0.0589	0.0418		
Score 16 or Higher	0.0448	0.0313		
Selected for Telephone Clinical Follow-Up, by K6 Score				
Score 0 to 3	23	74	0.006	0.020
Score 4 to 5	23	42	0.006	0.012
Score 6 to 7	29	37	0.008	0.010
Score 8 to 9	51	45	0.014	0.013
Score 10 to 11	60	45	0.017	0.013
Score 12 to 15	112	79	0.031	0.022
Score 16 or Higher	85	59	0.024	0.017
Total Selected for Telephone Clinical Follow-Up	382	382	0.106	0.106
Percent Agreeing to Clinical Follow-Up	0.86	0.86		
Percent Completing the Clinical Follow-Up Interview	0.76	0.76		
Completed Clinical Follow-Up Interview, by K6 Score	250	250	0.069	0.069
Score 0 to 3	15	48	0.004	0.013
Score 4 to 5	15	28	0.004	0.008
Score 6 to 7	19	24	0.005	0.007
Score 8 to 9	33	30	0.009	0.008
Score 10 to 11	39	30	0.011	0.008
Score 12 to 15	73	52	0.020	0.014
Score 16 or Higher	56	39	0.015	0.011

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

[1] The actual sampling rates were slightly higher than those shown in this table. In quarter 1, they were increased to account for the reduced 18 or older sample. In quarter 2, they were higher to compensate for a low clinical interview yield in quarter 1.

NOTE: The estimated distribution of interview respondents is based on the 2008 NSDUH. Estimated MHSS response rates are based on the 2008 Mental Health Surveillance Clinical Follow-Up Study.

Among NSDUH respondents aged 18 or older, 96 percent were eligible to be sampled for the MHSS clinical follow-up study (i.e., they completed the NSDUH main interview in English). At the end of the 2009 data collection, 87 percent of those selected for the follow-up agreed to participate, and 78 percent of those persons completed the clinical interview. The 2009 MHSS resulted in 521 completed clinical interviews. However, one case was excluded from the final dataset because of incomplete data (see Section 3.8.5 for further details). The final number of completed interviews in 2009 was 520. A summary of the 2009 MHSS clinical respondents by quarter is included in Table 3.7.

Table 3.7 2009 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	11,403	11,963	11,264	10,979	45,609
Unweighted K6 Distribution					
Score 0 to 3	0.45	0.45	0.45	0.45	0.45
Score 4 to 5	0.14	0.14	0.14	0.14	0.14
Score 6 to 7	0.10	0.10	0.10	0.11	0.10
Score 8 to 9	0.07	0.07	0.07	0.07	0.07
Score 10 to 11	0.06	0.05	0.05	0.06	0.06
Score 12 to 15	0.09	0.09	0.09	0.09	0.09
Score 16 or Higher	0.09	0.09	0.09	0.08	0.09
Eligible for MHSS	10,930	11,452	10,786	10,540	43,708
Eligibility Rate	*0.9585*	*0.9573*	*0.9576*	*0.9600*	*0.9583*
Selected for Telephone Clinical Follow-Up[1]	182	192	211	204	789
Zero Probability Cases	0	0	0	21	21
Agreed to Clinical Follow-Up	156	167	183	159	665
Percent Agreeing to Clinical Follow-Up (Including Zero Probability Cases)	*0.8571*	*0.8698*	*0.8673*	*0.7794*	*0.8428*
Percent Agreeing to Clinical Follow-Up (Excluding Zero Probability Cases)	*0.8571*	*0.8698*	*0.8673*	*0.8689*	*0.8659*
Completed Clinical Follow-Up Interviews	123	125	142	130	520
Clinical Follow-Up Interview Completion Rate	*0.7885*	*0.7485*	*0.7760*	*0.8176*	*0.7820*

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

[1] Includes cases assigned a zero probability of selection that would have been selected based on their K6 rates.

NOTE: The overall response rate to the MHSS clinical follow-up study should also consider the response rate to the main study. In 2009, the weighted overall response rate for the main NSDUH study was 66.79 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2009.

3.6 2010, 2011, and 2012 Sample Allocations

The 2010 sample was designed to yield 500 clinical interviews, and the 2011 and 2012 samples were designed to yield 1,500 clinical interviews. Because the 2010 through 2012 sample designs were similar, they are all described in this section.

The sample selection algorithms used in 2008 and 2009 led to some respondents having much greater weights than others, which resulted in large SEs of estimates. Because young persons aged 12 to 25 are oversampled in the NSDUH main study, the unweighted sample distribution by age in the main NSDUH sample does not reflect the distribution of the population by age. That is, a respondent aged 18 to 25 in NSDUH represents a smaller proportion of the population than a respondent aged 50 or older; therefore, the younger respondent has a much smaller weight. Because age was not a consideration in the sample selection algorithms in 2008 and 2009, this overrepresentation of young persons was maintained in the clinical sample; thus, the unequal weighting in the main NSDUH sample was maintained in the clinical sample.

Because this oversampling of young adults was not needed for the clinical sample's analytic goals, and the extremely unequal weights were adversely affecting the variance of the

survey estimates, a new selection algorithm was developed for the 2010 through 2012 clinical samples that effectively neutralized the effects of the NSDUH main interview oversampling. Subsamples of eligible respondents were selected for clinical follow-up with probabilities based on their K6 scores and WHODAS scores and on their age. Specifically, an age group equalization factor was used to adjust the K6- and WHODAS-based selection probabilities such that persons were selected for the clinical follow-up sample in accordance with the age distribution in the U.S. adult population. Essentially, this led to young persons in the main NSDUH sample being selected for the clinical follow-up study at a lower probability compared with persons in older age groups. This approach resulted in weights across age groups that were less variable.

Table 3.8 shows some of the age-related factors used to compute sampling rates in 2010, 2011, and 2012. For example, based on the 2008 population estimates and the 2010 planned sample, the average weighting for persons aged 50 or older was almost 10 times as large as the average weight for persons aged 18 to 25. (Smaller differences occurred for the intermediate age groups—26 to 34 and 35 to 49.) To compensate for this initial disparity in weights and to focus on persons aged 18 or older as a whole, sampling rates were set for persons aged 18 to 25, then increased for the other three age groups by adjusting the sampling rate using the final age-related weight equalization factor shown in Table 3.8. For each age group, the derived weight equalization factor is equal to the average weight for that age group divided by the average weight for the 18 to 25 age group. For example, the weight equalization factor in 2010 for the 26 to 34 age group equals 5,656/1,464 ≈ 3.8637 (see Table 3.8).[10] Because the average weight for persons aged 50 or older was so much higher than the other age groups, use of the derived weight equalization factors would have greatly increased the sampling rate for persons aged 50 or older. An adjusted set of factors that partially reduced the unequal weighting effects across age groups was specified instead. Rather than using a different age equalization factor for each age group, the adjusted equalization factors for the 35 to 49 and 50 or older age groups were set equal to the factor for the 26 to 34 age group.

The general sample allocation strategy was to find an allocation that provided a more precise estimate of all-adult SMI prevalence so that appropriate cut points (i.e., points in the SMI predicted probability continuum at which cases would be classified as SMI or not; see Chapter 7 for more details) could be established based on the MHSS clinical sample. To achieve this allocation, a total of 225 strata were defined based on the combination of 25 possible K6 scores (0 to 24)[11] and 9 possible WHODAS scores (0 to 8). The 225 strata were further split based on age group (18 to 25, 26 to 34, 35 to 49, and 50 or older), for a total of 900 allocation cells.

[10] The weight equalization factors were calculated using the average weights with decimals (not rounded average weights).

[11] In the prediction model, a recoded form of K6 score was used: Scores of 0 to 7 were recoded as 0, and all other scores had 7 subtracted from them to give a recoded total ranging from 0 to 17. These scores were reverse recoded to return to the original K6 scores that were used in the two-way matrix. This explains why the predicted probabilities of mental illness are all identical for K6 scores of 0 to 7.

Table 3.8 Age-Related Factors for the 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study

Year	Age	Population	Interview Respondents Aged 18 or Older	Average Weight	Initial Age-Related Weight Equalization Factor	Eligibility Factor[1] (%)	Response Rate Factor[2] (%)	Final Age-Related Weight Equalization Factor[3]
2010	18 to 25	32,938,184	22,500	1,464	1.0000	96.11	67.43	1.00000
	26 to 34	35,634,108	6,300	5,656	3.8637	93.20	66.87	4.01811
	35 to 49	64,198,531	9,700	6,618	3.8637	94.66	62.33	4.24452
	50 or Older	92,151,942	6,500	14,177	3.8637	96.79	58.72	4.40612
2011	18 to 25	33,579,988	22,500	1,492	1.0000	96.69	69.27	1.00000
	26 to 34	36,214,628	6,000	6,036	4.0442	93.66	62.18	4.65050
	35 to 49	63,166,074	9,000	7,018	4.0442	94.86	67.59	4.22450
	50 or Older	94,245,857	7,500	12,566	4.0442	96.73	65.48	4.27670
2012	18 to 25	34,072,349	22,500	1,514	1.0000	97.08	67.72	1.00000
	26 to 34	36,523,574	6,000	6,087	4.0198	93.03	71.89	3.95170
	35 to 49	62,042,733	9,000	6,894	4.0198	94.64	67.97	4.10880
	50 or Older	96,633,922	7,500	12,885	4.0198	96.76	60.00	4.55250

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

[1] Only NSDUH respondents who had completed the main interview in English were eligible for the clinical follow-up.

[2] Response rates shown are the product of the percentage agreeing to the clinical follow-up survey and the percentage of those who actually participated.

[3] The final age-related equalization factor is the initial weight equalization factor divided by the eligibility and response rate factors and then normalized.

NOTE: Population estimates were made using 2008 NSDUH data for the 2010 design, 2009 NSDUH data for the 2011 design, and 2010 NSDUH data for the 2012 design. Eligibility and response rate factors were computed using 2008 MHSS clinical data for the 2010 design, 2009 MHSS clinical data for the 2011 design, and 2010 MHSS clinical data for the 2012 design.

To determine the sampling rates by stratum (denoted as h) and age group, the proportionality factors ($r_{h,age}$) were calculated based on the estimated probability of SMI, the estimated eligibility and response rates, and the final age-related weight equalization factors. Statistical models were developed to estimate the probability of SMI based on the K6 and WHODAS scores, while eligibility rates, response rates, and weight equalization factors were defined based on the age groups. The proportionality factors were defined as follows:

$$r_{h,age} \propto \frac{\sqrt{P_h(1-P_h)}}{E_{age}RR_{age}} * F_{age},$$

where P_h refers to the predicted probability of SMI in stratum h, and F_{age}, E_{age}, and RR_{age} refer to the age-specific weight equalization factors, eligibility factors, and response rate factors, respectively. These proportionality factors then were multiplied by the projected sample counts and scaled to achieve the desired overall respondent sample (500 persons aged 18 or older in 2010 and 1,500 persons aged 18 or older in 2011 and 2012) and to obtain the stratum and age-specific sampling rates.

As an example, from the 2011 MHSS clinical data the predicted probability of SMI for a person with a K6 score of 10 and a WHODAS score of 6 was 0.1398. For the 18 to 25 age group, the proportionality factor then would be

$$r_{h,18-25} = \frac{\sqrt{0.1398(1-0.1398)}}{0.9669*0.6927} *1.000 = 0.5177.$$

The proportionality factors were then scaled in order to achieve an overall sample of 1,500 persons. A scaling factor of 0.0885 was expected to yield 1,500 completed interviews, so it was applied to each proportionality factor. Thus, the sampling rate for this stratum and age group was 0.5177 * 0.0885 = 0.0458.

Projected yields of positive cases based on the predicted probability of SMI and AMI broken out by age group and year are provided in Table 3.9. Table 6.1 in Chapter 6 presents the total number of Structured Clinical Interview for DSM-IV (SCID) respondents for each age group by year,[12] and Table 6.2 presents the total number of SCID respondents who were diagnosed with SMI or AMI by year in comparison.

In addition, Table 3.10 provides the 2010-2012 MHSS clinical sample allocations by K6 group, and Table 3.11 provides the 2010-2012 MHSS clinical sample allocations by WHODAS score. The 2010 probability sample of 500 clinical follow-up interviews was distributed across four calendar quarters with approximately 125 clinical follow-up interviews per quarter. The 2011 and 2012 samples of 1,500 clinical follow-up interviews were distributed across four calendar quarters with approximately 375 clinical follow-up interviews per quarter. Throughout the 2010, 2011, and 2012 surveys, the MHSS clinical sample was monitored, and the sampling

[12] See Chapter 1 for details on the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP or SCID).

Table 3.9 Projected Yields of Predicted Positive Cases, by Age Group: 2010 to 2012 MHSS Clinical Follow-Up

Year	Measure	18 to 25	26 to 34	35 to 49	50 or Older	18 or Older
2010	SMI	25	23	31	10	89
	AMI	67	59	80	33	239
	Total	**116**	**116**	**170**	**98**	**500**
2011	SMI	72	71	92	39	274
	AMI	194	184	226	118	721
	Total	**335**	**343**	**477**	**345**	**1,500**
2012	SMI	76	69	92	45	281
	AMI	195	174	224	126	720
	Total	**338**	**336**	**476**	**351**	**1,500**

AMI = any mental illness; K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

NOTE: Predicted probabilities of SMI and AMI were calculated for each of the 225 K6 by WHODAS strata using models based on 2008 Mental Health Surveillance Clinical Follow-Up Study data. These probabilities were applied to the anticipated MHSS Clinical Follow-Up Study distributions for 2010 through 2012 to obtain estimates at the age group level. Sample distributions were estimated using 2008 NSDUH data for the 2010 design, 2009 NSDUH data for the 2011 design, and 2010 NSDUH data for the 2012 design.

Table 3.10 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study Sample Allocations, by K6 Group

Year	K6 Group	Percent of Population[1]	Assumed SMI Rate (%)[2,3]	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
2010	0 to 3	53.28	0.91	162	0	0.00794
	4 to 5	14.22	1.20	51	0	0.00790
	6 to 7	9.28	1.73	38	0	0.00841
	8 to 9	6.06	2.98	34	0	0.01060
	10 to 11	4.74	5.01	35	1	0.01373
	12 to 15	6.55	12.65	78	18	0.01958
	16 or Higher	5.87	39.03	103	69	0.02620
	Total	**100.00**	**4.36**	**500**	**89**	
2011	0 to 3	53.21	0.92	491	0	0.02322
	4 to 5	13.78	1.21	148	0	0.02293
	6 to 7	9.41	1.61	116	0	0.02418
	8 to 9	5.96	2.65	96	0	0.03034
	10 to 11	4.64	5.32	101	3	0.03978
	12 to 15	6.99	12.49	233	50	0.05569
	16 or Higher	6.02	41.07	316	221	0.07751
	Total	**100.00**	**4.56**	**1,500**	**274**	
2012	0 to 3	53.54	0.92	497	0	0.02370
	4 to 5	13.67	1.21	145	0	0.02386
	6 to 7	9.08	1.58	114	0	0.02534
	8 to 9	6.26	2.86	98	0	0.03126
	10 to 11	4.41	5.51	99	4	0.04192
	12 to 15	6.71	12.28	225	52	0.05731
	16 or Higher	6.33	41.85	322	225	0.07980
	Total	**100.00**	**4.70**	**1,500**	**281**	

K6 = Kessler-6, a 6-item psychological distress scale; NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

[1] Source: 2008 NSDUH data for the 2010 design, 2009 NSDUH data for the 2011 design, and 2010 NSDUH data for the 2012 design.

[2] Source: 2008 NSDUH clinical follow-up study for the 2010 design, 2009 NSDUH clinical follow-up study for the 2011 design, and 2010 NSDUH clinical follow-up study for the 2012 design.

[3] To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across K6 scores. These rates are not the actual SMI rates that were used in the sample allocation.

Table 3.11 2010 to 2012 Mental Health Surveillance Clinical Follow-Up Study Sample Allocations, by WHODAS Score

Year	WHODAS Score	Percent of Population[1]	Assumed SMI Rate (%)[2,3]	Expected Sample Size	Expected SMI Count	Overall Sampling Rate
2010	0	74.80	1.09	248	0	0.00736
	1	7.11	2.71	39	1	0.01207
	2	5.30	4.26	36	1	0.01498
	3	3.14	8.01	32	2	0.02276
	4	2.57	11.80	29	7	0.02492
	5	2.04	19.04	29	11	0.03140
	6	1.74	31.80	31	21	0.04003
	7	1.32	41.36	25	18	0.04255
	8	1.98	54.53	31	28	0.03518
	Total	**100.00**	**4.36**	**500**	**89**	
2011	0	74.50	1.05	734	0	0.02188
	1	7.21	2.49	115	2	0.03529
	2	4.99	4.55	108	5	0.04829
	3	3.32	8.51	97	9	0.06485
	4	2.48	13.90	92	23	0.08205
	5	2.38	20.42	95	38	0.08876
	6	1.93	30.34	90	56	0.10363
	7	1.28	43.47	70	55	0.12286
	8	1.92	58.32	100	87	0.11535
	Total	**100.00**	**4.56**	**1,500**	**274**	
2012	0	74.26	1.06	738	0	0.02210
	1	7.41	2.61	116	3	0.03473
	2	4.95	4.67	104	4	0.04695
	3	3.37	7.68	93	8	0.06132
	4	2.61	13.45	89	22	0.07585
	5	2.05	21.26	89	36	0.09634
	6	1.83	32.44	87	57	0.10606
	7	1.44	42.75	75	56	0.11673
	8	2.10	58.81	108	95	0.11413
	Total	**100.00**	**4.70**	**1,500**	**281**	

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Scale.

[1] Source: 2008 NSDUH data for the 2010 design, 2009 NSDUH data for the 2011 design, and 2010 NSDUH data for the 2012 design.

[2] Source: 2008 MHSS clinical follow-up study for the 2010 design, 2009 MHSS clinical follow-up study for the 2011 design, and 2010 MHSS clinical follow-up study for the 2012 design.

[3] To compute assumed SMI rates, SMI estimates by K6 and WHODAS score were averaged (weighted) across WHODAS scores. These rates are not the actual SMI rates that were used in the sample allocation.

parameters were modified on an as-needed basis. Sampling rates were adjusted as needed to ensure that the targeted number of completions was achieved for each year. For the 2010 and 2011 clinical study, for the last 3 weeks in quarter 4,[13] the probability of selection of the NSDUH respondents for the clinical follow-up survey was set to zero (see Section 3.8.1 for further details). For the 2012 MHSS, the probability of selection of the NSDUH respondents for the clinical follow-up survey was not set to zero. Respondents were selected for clinical follow-up through December 20, 2012 (the last day of NSDUH main study data collection), and MHSS clinical data collection continued through January 17, 2013.

The 2010 MHSS clinical study resulted in 516 completed clinical interviews. Approximately 84 percent of selected persons agreed to participate, and 81 percent of those persons completed the MHSS clinical interview. A summary of the 2010 respondents by quarter is included in Table 3.12.

The 2011 MHSS clinical study resulted in 1,495 completed clinical interviews. An estimated 84 percent of selected persons agreed to participate, and 79 percent of those persons completed the MHSS clinical interview. A summary of the 2011 respondents by quarter is included in Table 3.13.

The 2012 MHSS clinical study resulted in 1,622 completed clinical interviews. An estimated 84 percent of selected persons agreed to participate, and 79 percent of those persons completed the MHSS clinical interview. A summary of the 2012 respondents by quarter is included in Table 3.14.

3.7 Creation of Variance Estimation Strata and Replicates

The nature of the NSDUH stratified, clustered sampling design requires that the design structure be taken into consideration when computing variances of survey estimates. Variance estimation strata and replicates specific to the clinical follow-up study were created to appropriately account for the sample design. To define the variance estimation strata for the MHSS clinical study, the 900 NSDUH main study variance strata were sorted, and groups of 9 adjacent main study variance strata were collapsed for each year of data collection. Thus, a total of 100 MHSS clinical interview variance strata were formed (and named MHVESTR). The number of variance strata formed was selected to optimize the number of degrees of freedom (*df*) available for analyses. The variance strata were formed for the purpose of producing national estimates only because subnational estimates are not recommended. It is recommended that the variable MHVESTR be used to produce SEs of estimates when using the combined 2008-2012 clinical data.

In this report, methodological analyses of the clinical interview data by individual year are included, even though the clinical interview data are intended to be used as a 5-year pooled dataset when conducting nonmethodological, substantive analyses. Prior to the completion of data collection, a preliminary strata variable (MHVSTR09) with only 50 strata was created and

[13] The MHSS clinical interview recruitment in both 2010 and 2011 ended on November 29th, and the MHSS clinical study and the NSDUH main study data collection ended on December 20th.

Table 3.12 2010 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,877	12,102	11,844	11,021	45,844
Unweighted K6 Distribution, by K6 Score					
Score 0 to 3	0.45	0.46	0.47	0.45	0.46
Score 4 to 5	0.14	0.14	0.14	0.13	0.14
Score 6 to 7	0.10	0.10	0.10	0.11	0.10
Score 8 to 9	0.07	0.07	0.07	0.07	0.07
Score 10 to 11	0.06	0.05	0.05	0.05	0.05
Score 12 to 15	0.09	0.08	0.09	0.09	0.09
Score 16 or Higher	0.09	0.09	0.09	0.09	0.09
Eligible for MHSS	10,446	11,608	11,341	10,563	43,958
Eligibility Rate	*0.9604*	*0.9592*	*0.9575*	*0.9584*	*0.9589*
Selected for Telephone Clinical Follow-Up[1]	190	246	175	157	768
Zero Probability Cases	0	0	0	4	4
Agreed to Clinical Follow-Up	163	198	146	133	640
Percent Agreeing to Clinical Follow-Up (Including Zero Probability Cases)	*0.8579*	*0.8049*	*0.8343*	*0.8471*	*0.8333*
Percent Agreeing to Clinical Follow-Up (Excluding Zero Probability Cases)	*0.8579*	*0.8049*	*0.8343*	*0.8693*	*0.8377*
Completed Clinical Follow-Up Interviews	132	157	115	112	516
Clinical Follow-Up Interview Completion Rate	*0.8098*	*0.7929*	*0.7877*	*0.8421*	*0.8063*

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = World Health Organization Disability Assessment Scale.

[1] Includes cases assigned a zero probability of selection that would have been selected based on their K6 and WHODAS scores.

NOTE: The overall response rate to the MHSS clinical follow-up study should also consider the response rate to the main study. In 2010, the weighted overall response rate for the main NSDUH study was 65.94 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2010.

used for the initial modeling and methodological work that analyzed data from each individual year. Because of the smaller yearly sample sizes, larger clinical interview variance strata were needed. Therefore, analyses in this report also used that preliminary version.

Two replicates (i.e., primary sampling units for variance estimation purposes) are defined within each NSDUH main study variance stratum. Each variance replicate consists of four segments, one for each quarter of data collection. The first replicate consists of those segments that are "phasing out" or will not be used in the next survey year. The second replicate consists of those segments that are "phasing in" or will be fielded again the following year, thus constituting the 50 percent overlap between survey years. The variance replicate assigned to the segment in which the MHSS clinical interview respondent was sampled was retained for the MHSS clinical study (and named MHVEREP). Further details about the formation of NSDUH main study

Table 3.13 2011 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,840	12,481	12,170	11,108	46,599
Unweighted K6 Distribution, by K6 Score					
Score 0 to 3	0.46	0.47	0.46	0.45	0.46
Score 4 to 5	0.14	0.13	0.13	0.14	0.14
Score 6 to 7	0.10	0.10	0.10	0.10	0.10
Score 8 to 9	0.07	0.07	0.07	0.07	0.07
Score 10 to 11	0.05	0.05	0.06	0.05	0.05
Score 12 to 15	0.09	0.08	0.09	0.09	0.09
Score 16 or Higher	0.09	0.09	0.09	0.09	0.09
Eligible for MHSS	10,392	11,974	11,665	10,709	44,740
Eligibility Rate	*0.9587*	*0.9594*	*0.9585*	*0.9641*	*0.9601*
Selected for Telephone Clinical Follow-Up[1]	543	672	531	531	2,277
Zero Probability Cases	15	0	0	26	41
Agreed to Clinical Follow-Up	450	561	449	421	1,881
Percent Agreeing to Clinical Follow-Up (Including Zero Probability Cases)	*0.8287*	*0.8348*	*0.8456*	*0.7928*	*0.8261*
Percent Agreeing to Clinical Follow-Up (Excluding Zero Probability Cases)	*0.8523*	*0.8348*	*0.8456*	*0.8337*	*0.8412*
Completed Clinical Follow-Up Interviews	363	436	359	337	1,495
Clinical Follow-Up Interview Completion Rate	*0.8067*	*0.7772*	*0.7996*	*0.8005*	*0.7948*

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = World Health Organization Disability Assessment Scale.

[1] Includes cases assigned a zero probability of selection that would have been selected based on their K6 and WHODAS scores.

NOTE: The overall response rate to the MHSS clinical follow-up study should also consider the response rate to the main study. In 2011, the weighted overall response rate for the main NSDUH study was 64.69 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2011.

variance strata and replicates can be found in the 2012 sample design report (Morton, Martin, Shook-Sa, Chromy, & Hirsch, 2013). When analyzing data from the MHSS clinical study, the variance estimation strata, replicates, and analysis weight must be taken into account using software capable of design-consistent estimation (e.g., SUDAAN®, STATA®).[14]

[14] SUDAAN® is used on NSDUH and stands for SUDAAN® Software for Statistical Analysis of Correlated Data (RTI International, 2012). STATA® is an integrated statistical software package that can be used for data analysis, data management, and graphics (http://www.stata.com/).

Table 3.14 2012 Mental Health Surveillance Clinical Follow-Up Study, Quarters 1 through 4 Summary

Design Parameter	Quarter 1	Quarter 2	Quarter 3	Quarter 4	Total
Interview Respondents Aged 18 or Older	10,894	12,144	11,997	10,801	45,836
Unweighted K6 Distribution, by K6 Score					
Score 0 to 3	0.45	0.46	0.46	0.46	0.46
Score 4 to 5	0.14	0.14	0.14	0.14	0.14
Score 6 to 7	0.10	0.10	0.10	0.10	0.10
Score 8 to 9	0.07	0.06	0.07	0.07	0.07
Score 10 to 11	0.05	0.06	0.06	0.05	0.06
Score 12 to 15	0.09	0.09	0.09	0.08	0.09
Score 16 or Higher	0.09	0.09	0.09	0.09	0.09
Eligible for MHSS	10,488	11,699	11,591	10,442	44,220
Eligibility Rate	*0.9627*	*0.9634*	*0.9662*	*0.9668*	*0.9647*
Selected for Telephone Clinical Follow-Up[1]	585	596	634	649	2,464
Zero Probability Cases	0	0	0	0	0
Agreed to Clinical Follow-Up	491	497	532	543	2,063
Percent Agreeing to Clinical Follow-Up (Including Zero Probability Cases)	*0.8393*	*0.8339*	*0.8391*	*0.8367*	*0.8373*
Percent Agreeing to Clinical Follow-Up (Excluding Zero Probability Cases)	*0.8393*	*0.8339*	*0.8391*	*0.8367*	*0.8373*
Completed Clinical Follow-Up Interviews	384	379	434	425	1,622
Clinical Follow-Up Interview Completion Rate	*0.7821*	*0.7626*	*0.8158*	*0.7827*	*0.7862*

K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = World Health Organization Disability Assessment Scale.

[1] Includes cases assigned a zero probability of selection that would have been selected based on their K6 and WHODAS scores.

NOTE: The overall response rate to the MHSS clinical follow-up study should also consider the response rate to the main study. In 2012, the weighted overall response rate for the main NSDUH study was 62.87 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2012.

3.8 Issues That Affected the Clinical Follow-Up Sample Design

3.8.1 Assigning NSDUH Clinical Interview-Eligible Cases Zero Selection Probability

The clinical interviews were conducted by telephone follow-up after the completion of the main NSDUH interview, so it was a logistical challenge for RTI to collect data from respondents selected for the MHSS clinical study late in quarter 4 because the NSDUH protocol is to suspend all data collection activities approximately 10 days prior to December 31st of the year. From 2008 to 2011, recruitment for the clinical interview was suspended early in order to complete data collection in time. This suspension of recruitment was accomplished by altering the clinical interview sampling algorithm in the main NSDUH interview such that respondents

who completed their main study interviews beyond the specified clinical interview cutoff date[15] would have a zero probability of selection for clinical follow-up. Respondents who completed their main NSDUH interview after the clinical interview cutoff point had no chance of being selected for the MHSS clinical interview, regardless of their K6 scores, WHODAS scores, or age. This suspension of recruitment for the clinical interview was not implemented in the 2012 MHSS. In 2012, recruitment for the clinical interviews continued through the end of main study data collection,[16] and selected cases continued to be contacted and interviewed after main study data collection concluded.

In addition to the cases that had no chance of selection because their main study interview was conducted after the clinical interview cutoff date, at the beginning of quarter 1 in the 2011 data collection, 15 NSDUH respondents who should have been selected for the clinical interview were inadvertently given a zero probability of selection for the MHSS clinical interview.

Eligible respondents who were given a zero probability of selection for the MHSS clinical interview study may have different mental health characteristics than persons who were given a chance of being selected. To avoid potential bias from the exclusion of zero probability cases, cases were identified that would have been selected for the MHSS clinical study based on their age (for 2010 to 2012 only), K6 scores, and WHODAS scores, and these zero probability cases were classified as nonrespondents rather than treated as not sampled cases. In 2008 and 2009, MHSS clinical interview selection status for the zero probability cases was determined by comparing the sampling rates for the cases' K6 rates with their random numbers to determine whether they would have been selected. For 2010 and 2011, the probability of selection took into account their K6 rates, WHODAS scores, and age group adjustments. All zero probability cases that would have been selected if they had not been given a zero probability of selection were treated as nonrespondents in the calculation of the MHSS clinical sample analysis weights (see Section 5.4 in Chapter 5 for more details on weighting adjustments for nonresponse).
The numbers of cases assigned a zero probability of selection that would otherwise have been selected for the clinical interview are shown in Tables 3.4, 3.7, 3.12, 3.13, and 3.14 for each year of MHSS clinical study data collection.

3.8.2 Noncoverage for Non-English Speakers

The target population for the MHSS clinical study excluded persons whose main study NSDUH interview was conducted in Spanish. Approximately 4 percent of the main NSDUH interviews are completed in Spanish each year. Among interviews with Hispanic respondents, 23 percent are conducted in Spanish.[17] The undercoverage of Spanish speakers was adjusted through a sample weighting procedure (see Section 5.3 in Chapter 5).

---

[15] Recruitment for the clinical interview was suspended with 4 weeks remaining in data collection in 2008 (November 21, 2008), with 5 weeks remaining in data collection in 2009 (November 16, 2009), and with 3 weeks remaining in data collection in 2010 and 2011 (November 29, 2010, and November 29, 2011).

[16] Both the main study data collection and the MHSS clinical interview recruitment were suspended on December 20, 2012. Clinical interviews were conducted through January 17, 2013.

[17] This percentage is based on 2012 NSDUH respondents aged 18 or older.

3.8.3 Gulf Coast Oversample

In 2011, a Gulf Coast Oversample (GCO) was included to measure the impact of the April 20, 2010, Deepwater Horizon oil spill on substance use, mental health, and the utilization of substance abuse and mental health services in the Gulf Coast region's areas of Louisiana, Mississippi, Alabama, and Florida. The 2011 main study sample was expanded by 2,000 completed interviews in specific counties and/or parishes in Alabama, Florida, Louisiana, and Mississippi.

With the additional sample in the Gulf Coast region, the main study sample shifted from approximately 1.3 percent in the affected area to 3.3 percent in this area. As a result, 56 MHSS clinical interviews (3.75 percent of the total) were completed in the affected area. The number of completed main study and clinical interviews in the Gulf Coast region by year are presented in Table 3.15.

Table 3.15 Completed NSDUH Main Interviews and MHSS Clinical Interviews in the GCO Region, by Year

Year	NSDUH Main Interviews	MHSS Clinical Interviews
2008	753	15
2009	783	5
2010	802	7
2011	2,313	56
2012	803	22
Total	**5,454**	**105**

GCO = Gulf Coast Oversample; MHSS=Mental Health Surveillance Study; NSDUH=National Survey on Drug Use and Health.

NOTE: The GCO region comprises the geographic areas in the GCO. The Gulf Coast Oil Spill Disaster Area includes the 32 counties and parishes in Louisiana, Mississippi, Alabama, and Florida that were most affected by the April 2010 oil spill: Iberia, Jefferson, Lafayette, Lafourche, Orleans, Plaquemines, St. Bernard, St. Martin, St. Mary, St. Tammany, Terrebonne, and Vermilion of Louisiana; George, Hancock, Harrison, Jackson, Pearl River, and Stone of Mississippi; Baldwin, Clarke, Escambia, Mobile, Monroe, and Washington of Alabama; and Bay, Escambia, Franklin, Gulf, Okaloosa, Santa Rosa, Wakulla, and Walton of Florida.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

3.8.4 Falsified Main NSDUH Interviews

At the beginning of quarter 4 of the 2011 NSDUH, it was discovered that an FI in Pennsylvania had been falsifying interviews throughout 2011. Data monitoring and field verification are part of a routine process to ensure the quality of NSDUH data. In a typical quarter when only a few falsified cases are identified, the falsified screening and interview cases are reworked, and the responsible FI is removed from the project. Screening and interview cases completed in earlier quarters by FIs found to have falsified data in the current quarter are typically not removed or reworked. Because of the large-scale falsification found in Pennsylvania, the procedure was changed to discard falsified data back to 2008 (when this FI began falsifying clinical records) in addition to reworking the quarter 4 cases.

All main study NSDUH screening interviews completed by this FI from 2008 through 2011 were deemed incomplete with unknown eligibility, and the corresponding interviews were dropped from the analysis files (Morton et al., 2012). Consequently, all clinical follow-up cases sampled from these falsified main study NSDUH interviews were recoded as main study incompletes. MHSS clinical data were reprocessed such that cases selected for the MHSS clinical interviews that were sampled from the falsified main study NSDUH interviews were no longer treated as sampled for the MHSS clinical interviews. MHSS clinical analysis weights were recalculated for 2008 through 2010 without these cases (i.e., the falsified cases were excluded from the nonresponse models and were considered as being neither respondents nor nonrespondents).

As shown in Table 3.16, the removal of falsified cases resulted in two MHSS clinical interviews being removed from the 2008 MHSS clinical follow-up. No completed clinical interviews were lost from the 2009 and 2010 MHSS clinical follow-ups, and the falsification was discovered prior to the processing of the 2011 MHSS clinical follow-up, so no reweighting was required.

Table 3.16 Falsified Cases Removed from the 2008-2010 Mental Health Surveillance Clinical Follow-Up Study

Type of Case Removed	2008 MHSS	2009 MHSS	2010 MHSS
Selected for Clinical Follow-Up	7	3	1
Agreed to Clinical Follow-Up	4	0	0
Completed Clinical Interviews	2	0	0

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH MHSS clinical sample, 2008-2012.

3.8.5 Respondents Excluded from 2008-2012 Adult Clinical Interview Data File

In 2008 and 2009, data from five respondents were excluded and treated as nonrespondents (see Section 5.4 in Chapter 5 for details on the nonresponse weighting adjustment for the SCID sample) in the final SCID dataset. These respondents were excluded either because the respondent had an extremely large weight or because responses on all K6 and WHODAS (or SDS) item scores were missing.

Respondents with low K6 total scores typically had relatively large weights because they were undersampled for the clinical follow-up. One case belonging to the 2008 WHODAS half sample with a large weight was designated as SMI positive by the SCID interview,[18] but the K6 and WHODAS total scores were zero, thus ensuring that this case would always be a false negative in the receiver operating characteristic (ROC) modeling analyses. The large weight (1.6 million for this respondent) had the effect of unduly influencing the ROC models developed in 2008, so this respondent was dropped from the dataset. Two respondents belonging to the 2008 SDS half sample also were removed for similar reasons (i.e., these cases had large weights

[18] See footnote 12.

and were designated as SMI positive by the SCID interview, which unduly influenced the ROC models). After the first two quarters of the data collection, a decision was made to allocate the clinical sample based on AMI rather than SMI in order to reduce the selection probabilities of respondents with extremely large weights while maintaining the efficiency of estimating SMI and AMI directly based on the clinical sample (see Section 3.5 for more information). In 2010 through 2012, in order to decrease the variability of the weights, the sample selection algorithm for the clinical sample was changed further so that the probability of selection was based on the respondent's age, K6 score, and WHODAS score. The new algorithm effectively neutralized the effects of the NSDUH main interview oversampling of young adults (see Section 3.6 for more information).

An additional 2008 respondent from the SDS half sample and a respondent from the 2009 MHSS clinical sample were removed because all of their item scores for the K6 components were missing. For respondents whose item scores for K6 and WHODAS (or SDS) were partially missing, a zero-assignment imputation method was used for both the item scores and total scores. Section 4.3 in Chapter 4 provides more details on how the missing values in the model predictors were treated.

4. Data Collection and Processing

4.1 Background

As stated in Chapter 1, the primary purpose of the Mental Health Surveillance Study (MHSS) is to provide clinical data for estimating a statistical model that predicts serious mental illness (SMI) based on data collected through the National Survey on Drug Use and Health (NSDUH). In Chapter 7, details are given on how this statistical model is developed and utilized for estimating SMI and other levels of mental illness. This chapter describes the data collection and process procedures used for the MHSS clinical study and the imputation procedures used on the NSDUH main study variables that are employed as predictors in the SMI prediction model.

4.2 SCID Data Collection

As discussed in Chapter 3, MHSS clinical interview respondents were selected and recruited for the follow-up clinical interview at the end of the NSDUH main study. Respondents agreeing to participate were asked to provide contact information, including first name, telephone number, an alternate telephone number if available, and the best days and times to call. The field interviewer (FI) entered the respondent's contact information into the NSDUH laptop and transmitted the data to RTI International. The next step was for a data collection manager to review the case and assign it to a clinical interviewer. The data collection manager took into consideration the respondent's time zone and best times and dates for contact, as well as the clinical interviewer's availability and time zone when assigning cases. The clinical interviewer made the first attempt to contact a respondent within 24 hours of receiving the assigned case to schedule an appointment for the interview. If the respondent was not reached on the first call attempt, subsequent calls were made during the respondent's preferred time frame. After the clinical interviewer made several attempts without success during the time period specified by the respondent, calls were made at times outside of the respondent's preferred time frame in an attempt to reach the respondent.

The clinical interviewer would contact the respondent by telephone to schedule and to conduct the interview. Clinical interviewers were instructed to verify that the correct respondent had been contacted, obtain informed consent, and obtain permission to record the interview using the script provided. Permission to record the interview was not a requirement to participate in the study; however, clinical interviewers were instructed to inquire about any refusal to record the interview and to address any respondent concerns, if possible. To ensure confidentiality and privacy, no identifying information was written in the clinical interview booklet or solicited during the recorded interview.

As discussed in Section 2.3 in Chapter 2, the follow-up interview was a modified version of the Structured Clinical Interview for the DSM-IV, or SCID (First et al., 2002). The SCID is a semistructured diagnostic interview used to assess psychiatric disorders according to the criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (DSM-IV; American Psychiatric Association, 1994). As a semistructured clinical interview, the SCID contains structured, standardized questions that are read verbatim and sequentially, then

combined with unstructured follow-up questions that the clinical interviewer tailors to the respondent based on clinical judgment and respondent reports. The SCID was administered over the telephone by clinical interviewers who had undergone extensive training with SCID experts from RTI and the SCID's developer from Columbia University.

Clinical interviewers were trained to use their clinical judgment to code each item based on the respondent's answers and presentation. Using the SCID booklet, each criterion symptom was coded as "1" (absent or false), "2" (subthreshold), "3" (threshold or true), or "?" (inadequate information). Each diagnosis assessed was coded as "1" (absent) or "3" (present). As discussed in Section 4.3.1, the clinical interviewer would also score the Global Assessment of Functioning (GAF) on a scale of 1 (persistent danger to self or others) to 100 (superior functioning, no symptoms). The GAF is based on the respondent's psychological, social, and occupational functioning. Clinical interviewers scored the GAF based on the respondent's worst functioning in the past 12 months. At the conclusion of the SCID, the clinical interviewer read the end of interview script, which thanked the respondent for his or her time and mentioned the possibility of needing to speak with a counselor. Clinical interviewers ensured that the respondent had the toll-free number to the National Lifeline Network should he or she wanted additional information about mental health services in his or her area. If the clinical interviewer felt that the respondent was at all in distress or potentially was a danger to himself or herself or others, a distressed respondent protocol (DRP) was followed.

In order to ensure that both the main NSDUH interview and the clinical interview covered the same reference period, each follow-up interview was completed over the telephone within 4 weeks of the date of the main NSDUH interview. After the call ended, the clinical interviewer completed the interviewer debriefing questions within the SCID and reviewed the booklet to ensure that the documentation was complete. As part of the interviewer debriefing section, clinical interviewers recorded information about problems encountered during the interview (i.e., distressed respondent, cognitive impairment, problems with privacy, comprehension, cooperation), stressors the respondents had experienced in the past 12 months, a diagnosis that needed further assessment, and the overall validity of the SCID data. If clinical interviewers were uncertain how to code a particular item or were feeling unsure about whether they had gathered enough information, they were instructed to consult with one of the clinical supervisors for his or her opinion within an hour of completing the interview. Within 48 hours after completion of the interview, the clinical interviewer uploaded the recorded audio file to the Web-based Case Management System (CMS) and edited and shipped the paper SCID to RTI for final editing and keying in-house. Once shipped, the clinical interviewer assigned the case a code of "80: Materials Shipment (SCID)" and entered the Federal Express tracking number in the notes. A data collection manager then tracked the shipment to ensure that the SCID arrived at RTI.

4.3 SCID Data Processing

When the SCID was received at RTI, it was assigned a status code of "81: SCID Received at Regent." The SCID was then reviewed by clinical supervisors, technical editors, and keying clerks, following the editing and keying process described the following sections.

4.3.1 Clinical Editing

When the SCID was delivered to the MHSS clinical supervisors, it was assigned a status code of "82: SCID Received at Editing." The clinical supervisors reviewed the data collected in all SCID booklets, item by item, comparing the notes provided by the clinical interviewer and the diagnostic rating and listening to the accompanying audio files as needed to ensure confidence in the data. The audio file (if recorded) was matched with the corresponding paper-and-pencil SCID booklet. For full reviews, both the audio recordings and the SCID booklet were reviewed in their entirety. Newly hired clinical interviewers received full reviews of their first two interviews until they demonstrated proficiency in administering all SCID modules. Of the remaining interviews, 10 percent were randomly selected for a full review of the audio recordings and SCID booklet in their entirety. For the remaining partial reviews, significant portions of the overview and selected modules were reviewed. Cases that presented a more complex clinical picture received a closer review (i.e., cases in which the constellation of symptoms, the clinical interviewer's notes, and/or what is known about the indicated disorder[s] warranted special attention, as determined by the clinical supervisor).

As part of the review, clinical supervisors ensured that the clinical interviewers followed project protocol for administering the interview and provided accurate and sufficient notes in the SCID booklet. The clinical supervisors also evaluated the clinical interviewers' clinical interviewing techniques, diagnostic skills, and the extent to which they captured the overall clinical picture relative to the symptoms, diagnoses, and GAF score. In the back of the SCID booklet, clinical supervisors rated the overall validity of the data and noted if there were any other disorders that needed further assessment, such as a disorder that was not assessed in the study or a disorder where more information was needed for diagnosis.

Quality control concerns arose when either a respondent did not provide complete and accurate information or the clinical interviewer did not adequately conceptualize the case or assess the symptoms. To ensure quality control throughout the clinical editing process, the clinical supervisors met to discuss complex cases and data that were difficult to code. Consensus ratings were used to code these data.

For each interview, a clinical supervisor completed a clinical editing form, noting strengths and areas for improvement for the clinical interviewer. The SCID editing forms provided a record of clinical editing, a format for individualized feedback, and a system for monitoring clinical interviewer performance. Completed SCID editing forms were filed in separate folders for each clinical interviewer in a locked filing cabinet. Once clinical editing was completed, the corresponding code of "88: SCID Delivered to Technical Editing" was entered into the CMS, and the SCID booklet was given to the technical editors.

Completed cases that one clinical supervisor considered to be of questionable validity were subject to an independent second review by at least one other clinical supervisor. To address the potential bias to exclude more complex or unusual cases and those with a difficult respondent, the level of difficulty and overall confidence in the case were considered prior to making a final disposition. If all of the clinical supervisors agreed, then the case was removed from the study. In these situations, a code of "90: SCID Not Keyed - Reason Unspecified"

(prior to 2010) or "90.X" (after 2010; "X" depending on the reason) was entered into the CMS. Any relevant feedback was given directly to the clinical interviewer.

4.3.2 Technical Editing

Once clinical editing was completed, technical editors reviewed each SCID to ensure that all variables met the criteria for consistency[19] and to verify that skip logic criteria were true. For example, both variables S1 and E28 in the SCID are used to code the same symptom (any panic attacks); therefore, consistency checks were done such that "IF S1 = 1, then E28 = 1" and to verify that conditions stated in the "skip logic" criteria were true. An example of the verification of the skip logic includes that for major depressive episode (MDE). If criterion A for past year MDE is not met and therefore absent, the remaining items to assess past year MDE are "skipped" and left blank, such that "IF A10 = 1, then A11-A17 = BLANK" (see the technical editing guide in Appendix H of the 2008-2012 MHSS operations report; CBHSQ, 2014a). Any inconsistencies or cases in which the conditions were not true were noted on a technical editing form. The technical editing form and its respective SCID were returned to the clinical supervisors for final review and correction. Once completed by the technical editors, a code of "89: Technical Editing Complete" was recorded in the CMS.

4.3.3 Keying

After technical editing was completed and any discrepancies were resolved, the data clerks keyed the SCID data into a computer and stored the keyed data in an electronic file. All codes that did not meet specifications were flagged by the machine edit program. A report of flagged cases was sent to the clinical supervisors indicating the line(s) where the errors were suspected. SCID booklets flagged with potential problems were retrieved from storage and compared with the computer-generated report. Those line(s) were reviewed again to ensure accuracy in both clinical and technical editing. Any errors were corrected and communicated to the dataset manager to be reset by hand if needed.

4.4 Missing Data

To ensure that complete cases were available for modeling mental illness, all missing values for item scores from the Kessler-6 (K6), World Health Organization Disability Assessment Schedule (WHODAS), and Sheehan Disability Scale (SDS) sets of variables were recoded with zeros, and the recoded item scores were used to calculate the total scores that were used as predictors in the SMI models (see Chapter 7 for more information on the SMI modeling). The item score variables that were recoded include the following:

- past month K6 variables (DSTNRV30, DSTHOP30, DSTRST30, DSTCHR30, DSTEFF30, DSTNGD30),
- past year K6 variables (DSTNRV12, DSTHOP12, DSTRST12, DSTCHR12, DSTEFF12, DSTNGD12), and

[19] The modified SCID questionnaire for the 2008-2012 MHSS clinical study is shown in Appendix C of the 2008-2012 MHSS operations report (Center for Behavioral Health Statistics and Quality [CBHSQ], 2014a).

- WHODAS variables (IMPREMEM, IMPCONCN, IMPGOUT, IMPGOUTM, IMPPEOP, IMPPEOPM, IMPSOC, IMPSOCM, IMPHHLD, IMPHHLDM, MPRESP, IMPRESPM, IMPWORK).

These recoded item scores were used to calculate the total scores of K6 (K6SCMON, K6SCYR, K6SCMAX, WSPDSC2), WHODAS (WHODASC2, WHODASC3), and SDS (SDSSC2, SDSSC3). These total score variables are available for all adults on both the annual NSDUH public use file (CBHSQ, Substance Abuse and Mental Health Services Administration [SAMHSA], 2013) and the NSDUH 2008-2012 adult clinical interview data file (CBHSQ, SAMHSA, 2014b). See Appendices A, B, and C for more details on the K6, WHODAS, and SDS item score and total score variables.

Table 4.1 displays the unweighted numbers and percentages of the adult respondents in the 2008 to 2012 NSDUH main study who had at least one of the six past month K6 item scores missing, had at least one of the WHODAS or SDS item scores missing, and had their WHODAS or SDS questions skipped out because the sum of all K6 item scores was zero. These cases had their missing values set to zero.

An evaluation comparing the zero-assignment imputation method (i.e. assigning all the missing values with zeros) and weighted sequential hot deck imputation method resulted in no differences in estimates for SMI and any mental illness (AMI) (CBHSQ, in press b). Therefore, the recoded K6 and WHODAS total scores on the data file were used in the SMI prediction models.

Two other variables from the NSDUH main study—having past year major depression episode (AMDEYR) and having past year serious suicidal thoughts (MHSUITHK)—also had their missing values recoded as zeros: MHSUITHK was recoded as MHSUTK_U, and AMDEYR was recoded as AMDEY2_U. The recoded variables (MHSUTK_U and AMDEY2_U) were used in the 2012 revised models (see Section 7.3.4 in Chapter 7 for more details) as predictors for the mental illness status of NSDUH respondents. These recoded variables are also available for all adults on both the annual NSDUH public use file (CBHSQ, SAMHSA, 2013) and the NSDUH 2008-2012 adult clinical interview data file (CBHSQ, SAMHSA, 2014b). Table 4.2 reports the unweighted missing numbers and percentages of these two variables. The missing rates of both variables were all less than 1 percent for the 5 successive years. From 2008 through 2012, the prevalence rates of past year MDE were always less than 7 percent, and the rates of past year serious suicidal thoughts were always less than 4 percent. Therefore, assigning missing values with zeros will be very close to any other statistical imputation methods when estimating for SMI and AMI.

Table 4.1 Unweighted Missing Numbers and Percentages of K6, WHODAS, and SDS Item Scores: 2008-2012 All-Adult NSDUH Main Study

Variable	At Least One Item Was Missing		All Items Were Missing	
	Number	Percent (%)	Number	Percent (%)
K6 Item Scores				
2008-2012	2,498	1.09	456	0.20
2008A	237	1.05	45	0.20
2008B	249	1.08	46	0.20
2009	470	1.03	83	0.18
2010	525	1.15	98	0.21
2011	517	1.11	92	0.20
2012	500	1.09	92	0.20
WHODAS/SDS Item Scores, Skipped				
2008A-2012 (WHODAS)	-	-	37,296	18.06
2008A (WHODAS)	-	-	3,903	17.25
2008B (SDS)	-	-	3,922	17.02
2009 (WHODAS)	-	-	8,053	17.66
2010 (WHODAS)	-	-	8,313	18.13
2011 (WHODAS)	-	-	8,658	18.58
2012 (WHODAS)	-	-	8,369	18.26
WHODAS/SDS Item Scores, Non-Skipped				
2008A-2012 (WHODAS)	2,095	1.01	400	0.19
2008A (WHODAS)	221	0.98	50	0.22
2008B (SDS)	307	1.33	123	0.53
2009 (WHODAS)	482	1.06	79	0.17
2010 (WHODAS)	487	1.06	100	0.22
2011 (WHODAS)	472	1.01	84	0.18
2012 (WHODAS)	433	0.94	87	0.19

2008A = 2008 sample A; 2008B = 2008 sample B; - (i.e., a hyphen) = not applicable; K6 = Kessler-6, a 6-item psychological distress scale; NSDUH=National Survey on Drug Use and Health; SDS=Sheehan Disability Scale; WHODAS = World Health Organization Disability Assessment Scale.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 4.2 Unweighted Missing Numbers and Percentages of MDE and Suicidal Thoughts: 2008-2012 All-Adult NSDUH Main Study

Variable	Number	Percent (%)
Past Year MDE		
2008-2012	1,847	0.80
2008A	167	0.74
2008B	190	0.82
2009	353	0.77
2010	385	0.84
2011	377	0.81
2012	375	0.82
Past Year Suicidal Thoughts		
2008-2012	805	0.35
2008A	83	0.37
2008B	72	0.31
2009	147	0.32
2010	163	0.36
2011	162	0.35
2012	178	0.39

2008A = 2008 sample A; 2008B = 2008 sample B; MDE = major depressive episode; NSDUH=National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

5. Sample Weighting

5.1 Background

The principal purpose of the subsample of clinical data collected as part of the Mental Health Surveillance Study (MHSS) was for use in estimating the prevalence of serious mental illness (SMI) among adults aged 18 or older. With the clinical data, a statistical model was developed in 2008 and revised in 2012 to predict whether an adult respondent to the main National Survey on Drug Use and Health (NSDUH) interview had SMI based on his or her main NSDUH interview responses. Sections 7.2 and 7.3 in Chapter 7 provide the details of the original 2008 SMI prediction model and the 2012 revised model. Although the weights for the clinical sample were created primarily for modeling SMI, another useful application was in providing nationally representative direct estimates of mental disorders.

The weights used in the 2008 modeling process were developed using fairly simple methods because of the relatively small sample size in a single year (for more details, see Chapter 4 in Liao et al., 2012). By contrast, with the availability of approximately 5,500 clinical interview cases accumulated from 2008 to 2012, more sophisticated weighting procedures could be developed. These procedures attempted to correct for two types of potential biases: (a) coverage biases resulting from not including in the clinical sample adults who responded to the main NSDUH interview in Spanish and (b) nonresponse biases resulting either from adults selected for the clinical follow-up study refusing to be recontacted or from those initially willing to be recontacted not providing useful clinical data (usually due to noncontact). Research has shown that the causes of these two types of nonresponse can be very different. The weighting adjustment procedures also incorporated techniques that increased the statistical efficiency (decreased the standard errors [SEs]) of direct estimates of mental disorders and were intended to improve the efficiency of SMI modeling as well.

This chapter describes how the final weights were created for use in the 2012 model revision process and for use in creating direct estimations based on the combined 5-year clinical sample from 2008 to 2012. The weights incorporated improved coverage, nonresponse, and poststratification adjustments that were performed separately for each year. The two 2008 clinical samples, one for respondents who were given World Health Organization Disability Assessment Schedule (WHODAS) questions in the NSDUH main interview (the 2008A sample) and one for respondents who were given Sheehan Disability Scale (SDS) questions (the 2008B sample), were treated as if they represented separate years (2008A and 2008B). The adjusted weights for each year were then combined using scaling factors for analyzing either the combined 2008-2012 or 2008A-2012 data.

Section 5.2 presents an overview of the revised weighting adjustments. Section 5.3 first provides comparisons on items from the main NSDUH interview between Hispanic respondents interviewed in English and Hispanic respondents interviewed in Spanish, then develops an

adjustment for this source of undercoverage.[20] Section 5.4 provides comparisons between respondents and nonrespondents at two different phases of nonresponse to the clinical interview. It then develops separate nonresponse adjustments for each phase. Section 5.5 first discusses the poststratification adjustment, then describes the scaling factors applied to the weights in order to increase the statistical efficiency when analyzing the combined 5-year data. Section 5.6 provides assessments of the overall impact of the adjustments, while Section 5.7 summarizes the new weighting process and its impact.

5.2 Components of the MHSS Analysis Weights

The weights for the 2008 modeling process were based on four adjustment factors: (1) the respondent's NSDUH adult-level analysis weight, (2) the inverse of the probability that the respondent was selected for the clinical interview, (3) a nonresponse adjustment, and (4) a poststratification adjustment. The NSDUH adult-level analysis weight, ANALWT, has 15 weight components. Each weight component represented either the selection probability at each selection stage or an adjustment on nonresponse, poststratification, or extreme weights. See Chen et al. (2014) for further details about the construction of NSDUH's adult-level analysis weight.

Out of necessity, weighting adjustments for the 2008 clinical data were based on a relatively small sample of roughly 1,500 clinical interviews (around 750 for each half sample). Subsequent years of data collection from 2008 to 2012 have resulted in the collection of about 5,500 clinical interviews overall. The larger sample size permitted a more detailed evaluation of the potential for bias due to undercoverage and nonresponse and the development of adjustments that can account for the specific undercoverage of Hispanics due to the clinical follow-up study only being conducted in English and the two phases of nonresponse in the clinical follow-up (i.e., at the end of the NSDUH main interview and during the attempt to administer the clinical interview).

The revised weighting adjustments incorporated three features designed to further reduce nonresponse bias and undercoverage. First, a separate coverage adjustment was added in order to focus on the undercoverage of Hispanics who completed the main NSDUH interview in Spanish, (see footnote 20; Section 3.8.2 provides more details on the noncoverage of non-English speakers). [21] Hispanic respondents who were eligible for the clinical follow-up had their weights adjusted to compensate for those who completed the interview in Spanish. This was done by forcing the total of the adjusted weights among Hispanics to equal the total of the main NSDUH Hispanic weights, both overall and within categories shown to be related to whether a Hispanic NSDUH respondent completed the main NSDUH interview in English. An example of such a category is adults with less than a high school education.

---

[20] The NSDUH main sampling weights are adjusted for undercoverage. The sole source of undercoverage in the clinical follow-up is the removal of adults responding to the main NSDUH interview in Spanish from the target population for clinical subsampling.

[21] By contrast, the coverage adjustment in the original weighting (i.e., the weighting used with the-2008 modeling process) occurred in the poststratification step. Although that step did correct for the general undercoverage of Hispanic adults, it did not try to account for the characteristics of those Hispanics who completed the main NSDUH interview in Spanish.

The second feature of the new weighting scheme employed two steps in nonresponse adjustment rather than a single step. The two steps corresponded with the two distinct phases of nonresponse in the clinical follow-up study: (1) initial nonresponse resulting from failing to agree to a clinical follow-up (see Section 5.4.1 for more details), and (2) final nonresponse from not completing the clinical follow-up after having agreed to do it earlier (see Section 5.4.1 for more details). Weights were first adjusted for initial nonresponse using variables found to be correlated with initial refusal at the end of the main NSDUH interview and expected to be related to key outcome measures of the Structured Clinical Interview for DSM-IV (SCID).[22] After this first step, the weights were adjusted again for final nonresponse using variables found to be correlated with final nonresponse and expected to be related to SCID measures.

The first three weighting adjustments were designed to remove biases due to undercoverage or nonresponse. The next weighting adjustment was a poststratification adjustment that was implemented to increase the statistical efficiency of estimators computed using the final adjusted weight. This adjustment forced the sum of the final weights for the clinical data to equal the sum of the final weights for the adult NSDUH main interview sample for a set of variables and variable interactions that predicted SMI.

Overall, the final revised weight (MHFNLWGT[23]) was the product of seven factors:

1. the respondent's NSDUH personal-level analysis weight (ANALWT);
2. a coverage adjustment to account for NSDUH respondents who completed that survey in Spanish;
3. the inverse of the probability that the respondent was selected for the clinical interview after completing the main NSDUH interview;
4. a nonresponse adjustment to account for those who, immediately after completing the main NSDUH interview, refused to be recontacted for a clinical interview;
5. a second nonresponse adjustment to account mostly for those who agreed to be recontacted but were unavailable for the clinical interview (this included a few who agreed to be recontacted for the clinical interview, but refused to respond when recontacted);
6. a poststratification or population weighting adjustment using control totals from the main NSDUH interview; and
7. a scaling factor used to combine clinical data across years efficiently.

Factors 2, 4, and 5 adjusted the weights to remove potential biases due to nonrandom exclusions and nonresponse. Factor 2 compensated for the exclusion of adults responding to the NSDUH main survey in Spanish from the target population for the clinical sample, while factors 4 and 5 compensated for different types of nonresponse to the clinical study. The weighting

[22] See Chapter 1 for details on the Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP or SCID) (First et al., 2002).

[23] This is the weight for the combined 2008-2012 clinical data. For the data excluding the 2008B sample, the final weight is MHFAAWGT.

adjustments were arranged in the same sequence as the corresponding rounds of exclusions, subsampling, and nonresponse. Beginning with the adult respondents to the NSDUH main survey (factor 1), adults responding in Spanish were first removed (causing the compensation adjustment in factor 2). This was followed by the probability selection of the clinical sample (captured in factor 3), the refusal of some sampled adults to be recontacted (factor 4), and so forth.

Table 5.1 provides a reference for the variable names given to various adjusted weights and adjustment factors used with the clinical data in the NSDUH 2008-2012 adult clinical interview data file and some mathematical denotations that are used in this report. The weighting adjustments in steps 4, 5, and 6 of the table were each composed of two adjustment factors, an initial trimming of a few large weights followed by a weighting adjustment that compensated for the trimming as it accomplished its designed task. Details on these steps are contained in Sections 5.4.3, 5.4.5, and 5.5.1, respectively, while step 2 is the focus on the next section.

5.3 Adjustment of Undercoverage of Hispanics

Because the clinical follow-up interview was conducted in English, adults who completed the main NSDUH interview in Spanish had no chance of being selected for the clinical interview. To the extent that differences exist on the main NSDUH interview measures between those who completed the main NSDUH interview in Spanish rather than in English, the fact that respondents in the clinical follow-up must have completed the NSDUH main interview in English could result in the estimates based on the clinical sample not being fully representative of the U.S. general population. The social stigma associated with mental health issues and alcohol and other substance use has been known to lead to underreporting of such sensitive behaviors among Hispanics, relative to non-Hispanics (e.g., Clark & Hill, 1991; Greenfield & Kaskutas, 1998; Maddahian, Newcomb, & Bentler, 1988; Silva de Crane & Spielberger, 1981). Additionally, the language of survey administration has been found to cue cultural values and norms among bilingual and bicultural respondents (Ross, Xun, & Wilson, 2002; Trafimow, Silverman, Mei-Tai Fan, & Shui Fun Law, 1997; Triandis, Davis, Vassiliou, & Nassiakou, 1965), suggesting that Hispanic respondents may answer culturally sensitive questions differently depending on whether they are asked in Spanish or in English.

Table 5.2 presents the number of Hispanic respondents in the adult NSDUH main study along with the numbers and percentages of those interviewed in Spanish by year of data collection. Although only around 4 percent of the NSDUH adult interviews between 2008 and 2011 were conducted in Spanish, over 25 percent of the adult Hispanic NSDUH respondents were interviewed in Spanish (roughly 35 percent when the numbers are weighted).[24]

[24] Because less than 1 percent of the adults who completed a main NSDUH interview in Spanish were not Hispanic, only the analysis weights of Hispanics were adjusted to compensate for NSDUH respondents who completed the survey in Spanish.

Table 5.1 Adjusted Weights and Adjustment Factors in the NSDUH 2008-2012 Adult Clinical Interview Data File

Step	Adjusted Weight	Initial Weight(s)	Adjustment/ Weighting Factor	Reason for Adjustment
1	ANALWT_A (w_k)	ANALWT	N/A	For 2009-2012, ANALWT_A = ANALWT; for 2008A and 2008B, ANALWT_A is used when each half sample is treated as a separate year.
2	ANALWT_E[1]	ANALWT_A	MHADJ_1 (A_{1k})	Compensates for NSDUH respondents who completed the NSDUH main interview in Spanish.
3	MHDSNWT[1] (F_k)	ANALWT_E[1]	MHWT1	Incorporates the clinical sample selection probabilities conditional on the eligible NSDUH respondents.
4	MHWTNR1[1]	MHDSNWT[1]	MHADJ_2a (A_{2ak}) MHADJ_2b (A_{2bk})	Compensates for selected individuals who do not agree to be recontacted for the follow-up clinical interview; MHADJ_2a is the trimming factor; MHADJ_2b is the adjustment factor.
5	MHWTNR2[1]	MHWTNR1[1]	MHADJ_3a (A_{3ak}) MHADJ_3b (A_{3bk})	Compensates for the remaining nonrespondents who initially agreed to be recontacted, but were unable to participate in the clinical interview due to other factors. MHADJ_3a is the trimming factor; MHADJ_3b is the adjustment factor.
6	MHNEWWGT	MHWTNR2[1]	MHADJ_4a (A_{4ak}) MHADJ_4b (A_{4bk})	Forces the sum of weights for the clinical data to match the sum of weights for the adult NSDUH main interview sample.
7	MHFNLWGT	MHNEWWGT	Scaling Factors for the 2008-2012 data	Rescales weights by year for getting SCID-based estimates of the 2008-2012 data.
7a	MHFAAWGT	MHNEWWGT	Scaling Factors for the 2008A-2012 data	Rescales weights by year for getting SCID-based estimates of the 2008A-2012 (WHODAS) data.

N/A = not applicable; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV; WHODAS = World Health Organization Disability Assessment Schedule.

NOTE: MHFNLWGT = ANALWT_A * MHADJ_1 * MHWT1 * MHADJ_2a * MHADJ_2b * MHADJ_3a * MHADJ_3b * MHADJ_4a * MHADJ_4b * Scaling Factors for the 2008-2012 data.

[1] These variables are not available in the NSDUH 2008-2012 adult clinical interview data file, but they can be derived by multiplying ANALWT_A with related adjustment factors. For example, ANALWT_E = ANALWT_A * MHADJ_1.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 5.2 Hispanic Respondents, by Language Version: 2008-2011 All-Adult NSDUH Main Study

	All Hispanic Respondents	Hispanic Respondents Interviewed in Spanish	
Year/Sample	**Count**	**Count**	**Weighted Percent**
2008A	3,369	1,024	34.9
2008B	3,392	1,055	36.8
2009	6,861	1,899	34.6
2010	6,889	1,883	34.5
2011	6,960	1,865	35.7
Combined	**27,471**	**7,726**	**35.2**

NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2011.

5.3.1 Descriptive Characteristics of Hispanics Responding in English and Spanish

To better understand the consequences of excluding NSDUH respondents interviewed in Spanish for the clinical follow-up study, the possibility for coverage bias is examined by first comparing responses to the NSDUH mental health and substance use measures among Hispanics responding to NSDUH in English and those responding in Spanish. Differences in demographic characteristics and responses to mental health and substance use measures were examined between Hispanic respondents interviewed in Spanish and English from 2008 to 2011 (see Tables D1a to D1c in Appendix D). A small number (30) of non-Hispanic adult respondents were also interviewed in Spanish in the 2008 to 2011 NSDUHs. Those respondents were excluded from these language comparisons. The weights used for the comparisons were the NSDUH analysis weights (ANALWT_A).

Hispanic respondents interviewed in English were significantly different from Hispanic respondents interviewed in Spanish on a number of demographic characteristics (see Table D1a):

- Hispanics interviewed in Spanish tended to be older than those interviewed in English. Only 10.5 percent of those interviewed in Spanish were between the ages of 18 and 25, while 25.2 percent of those interviewed in English were similarly aged.
- Hispanics interviewed in Spanish were less educated than those interviewed in English. For example, 62.5 percent of those interviewed in Spanish reported having less than a high school education as compared with 21.3 percent of Hispanics interviewed in English.
- Hispanics interviewed in Spanish had significantly higher representation in the lowest income brackets. For example, 37.7 percent of those interviewed in Spanish reported an income of less than $20,000 as compared with 20.0 percent of those interviewed in English.
- Those interviewed in English were more likely to be born in the United States relative to those interviewed in Spanish (69.4 vs. 4.1 percent).
- A significantly higher percentage of Hispanics interviewed in English reported having health insurance coverage (71.1 percent) compared with those interviewed in Spanish (45.2 percent).

- A higher percentage of Hispanic respondents interviewed in English reported having no difficulty understanding the main NSDUH interview (88.3 percent) relative to those interviewed in Spanish (78.7 percent).
- Consistent with the literature on mental health issues, a significantly lower percentage of Hispanics interviewed in Spanish reported having been diagnosed with depression in their lifetime (5.3 percent) compared with Hispanics who were interviewed in English (8.9 percent).
- Spanish-language interviews with Hispanics yielded significantly lower estimates on mental health measures compared with English-language interviews with Hispanics, including Kessler-6 (K6) and WHODAS scores, lifetime and past year major depressive episode (MDE; i.e., depression), mental health treatment, suicidal thoughts, and treatment for attempted suicide (Table D1b). Specifically, the average K6 score among Hispanics responding in English was 5.1, while the mean for those responding in Spanish was 3.3. Similarly, past year MDE was 6.4 percent among Hispanics responding in English and 3.3 percent among those interviewed in Spanish.[25]
- Hispanics interviewed in English showed higher lifetime and past year prevalence rates than those interviewed in Spanish (Table D1c). Specifically, 50.1 percent of Hispanics responding in English reported any illicit lifetime drug use, while only 17.0 percent of Hispanics interviewed in Spanish reported such use.[26]

These results suggest that estimates related to mental health or substance use that are based on the clinical sample may be biased because of the exclusion of Hispanics who were interviewed in Spanish. The next section describes a method of adjusting for this source of undercoverage.

5.3.2 Development of the Coverage Adjustment to Compensate for Adults Interviewed in Spanish

The NSDUH weight (ANALWT) for an Hispanic adult respondent k to the clinical follow-up interview was adjusted yearly by a factor (MHADJ_1) of the form:

$$A_{1k} = \frac{1 + \exp(\mathbf{g}^T \mathbf{z}_k)}{1 + \exp(\mathbf{g}^T \mathbf{z}_k)/7}. \tag{5.1}$$

The adjustment factor in equation (5.1) comes from the implicit estimation of the probability that an Hispanic NSDUH respondent was interviewed in English based on his or her characteristics. Those characteristics are put into mathematical form by the components of the

[25] Further language comparisons focused on NSDUH's mental health measures are available in Table D1b of Appendix D.

[26] Language comparisons focused on NSDUH's more specific substance use measures are available in Table D1c of Appendix D.

vector $\mathbf{z}_k$. The estimated probability for respondent k with the characteristics captured in $\mathbf{z}_k$ is $1/A_{1k}$. Equation (5.1) implies that this estimated probability cannot be less than 1/7.

Using combined NSDUH data from 2008 to 2011, a weighted (by ANALWT_A) logistic regression analysis among Hispanics was used to pare down the choice of candidates for $\mathbf{z}_k$ from a much larger potential list.[27] The dependent variable in this analysis was whether a Hispanic NSDUH respondent completed the survey in English. Predictor variables in the regression analysis were selected as components of $\mathbf{z}_k$ when they were significant at an alpha level of 0.02.

The components of $\mathbf{z}_k$ finally selected for equation (5.1) were

- a categorical age group variable (18 to 25, 26 to 34, and 35 or older);
- a categorical education-level variable (less than high school, high school graduate, other);
- a categorical variable for the number of years the respondent lived in the United States (born in the United States, else < 5 years, else < 10 years, else);
- an indicator of whether the respondent was alone when answering the main NSDUH interview; and
- the respondent's NSDUH K6 score.[28]

Using the WTADJUST procedure in SUDAAN 11.0 (RTI International, 2012),[29] a parameter vector $\mathbf{g}$ in equation (5.1) was found yearly that satisfied the calibration equation:

$$\sum_S w_k \mathbf{z}_k = \sum_S w_k A_{1k} \mathbf{z}_k, \tag{5.2}$$

where S was the yearly NSDUH adult respondent sample, and w_k was the NSDUH analysis weight for respondent k. When k completed the main NSDUH interview in Spanish, A_{1k} was defined to be 0.[30] This allowed both sides of the equality in equation (5.2) to be summed across the entire NSDUH respondent sample. Equations (5.1) and (5.2) forced the weighted eligible-adult NSDUH sample for a clinical interview in a year to look like the full adult NSDUH main interview sample with respect to the components of the vector $\mathbf{z}_k$. Note that a different $\mathbf{g}$ is estimated every year using the same vector of components.

---

[27] These candidates were gender, five age group levels, age (continuous), four education levels, years lived in the United States (four categories), privacy when answering NSDUH, WHODAS score (continuous), adjusted WHODAS score (continuous), K6 score (continuous), adjusted K6 score (continuous), NSDUH lifetime MDE (binary), NSDUH past year MDE (binary), and suicidal thoughts (binary). See Chapter 2 for a discussion of the alternative K6 and WHODAS scores. For the 2008B sample, WHODAS and alternative WHODAS scores were replaced by an SDS measure. Hispanic respondents interviewed in English were found to be significantly different from Hispanic respondents interviewed in Spanish on these candidate variables.

[28] K6SCMAX is the respondent's K6 score (0 to 24) for his or her worst month in the past year.

[29] The generalized exponential model (GEM) macro, which had been used in adjusting the main NSDUH's person-level weights (see Chen et al., 2014), could have been employed in place of WTADJUST to produce the same results.

[30] This treats undercoverage like nonresponse and sets the lower bound at 1 and the upper bound at 7.

The **g** that satisfies equation (5.2) is a consistent estimator for **γ** when the probability (i.e., propensity) that a Hispanic *k* responds in English has the following truncated logistic form:

$$\rho_k = \frac{1+\exp(\boldsymbol{\gamma}^T \mathbf{z}_k)/7}{1+\exp(\boldsymbol{\gamma}^T \mathbf{z}_k)}.$$

Were 7 replaced by infinity and γ by −γ in the above equation, ρk would have the standard logistic form. Using the calibration equation to find **g** is an alternative method to (weighted) maximum likelihood.

To show how the coverage adjustment modified the NSDUH adult-level weights so that the Hispanics who responded to the NSDUH main interview in Spanish (and were excluded from the clinical sample) were represented by Hispanics with otherwise similar characteristics who responded to the NSDUH main interview in English (and were included in the clinical sample), it is helpful to divide adult Hispanic respondents to the NSDUH main survey into five *propensity strata* based on each respondent's implicitly estimated propensity (i.e., probability) of responding in English.[31]

$$p_k = \frac{1+\exp(\mathbf{g}^T \mathbf{z}_k)/7}{1+\exp(\mathbf{g}^T \mathbf{z}_k)}.$$

The five propensity strata are the adults with the highest 60 percent of p_k values, next 20 percent, next 10 percent, next 5 percent, and lowest 5 percent. Table 5.3 displays the ratio of the weighted population totals for adults after the coverage adjustment to before the coverage adjustment in each propensity stratum. The ratio is close to 1 for the highest response propensity stratum with 60 percent (11,847 divided by 19,745) of the unweighted sample of Hispanics who responded to the NSDUH main interview in English, but it is over 6 for the lowest response propensity stratum. This means that the weights for those adults most likely to have answered the NSDUH main interview in Spanish increased roughly 600 percent on average because of the coverage adjustment, while they increased hardly at all for over 80 percent of the adult Hispanics. Compared with the higher propensity strata, those in the lowest propensity strata tended to have characteristics more similar to Hispanics who responded to the main NSDUH interview in Spanish (i.e., those excluded from the clinical interview). Hence, increasing the weights for the lowest propensity stratum compensates for the undercoverage of the excluded Hispanics.

5.4 Nonresponse Adjustment

5.4.1 Components of Nonresponse in the MHSS Clinical Follow-Up

Nonresponse to the MHSS clinical interview consists primarily of two components that reflect the two-stage process of participation in the survey. First, a respondent selected from the

[31] In principle, all adult Hispanics have a probability-of-responding-in-English of this form, although only those who actually responded to the NSDUH main interview in English were assigned here to propensity strata.

main NSDUH interview to participate in the MHSS clinical follow-up was asked whether or not he or she would like to participate in the clinical interview, as Exhibit 5.1 shows.

Table 5.3 Summary of Coverage Adjustments for Excluding Spanish-Language Version Respondents, by Propensity Strata: 2008 to 2011 NSDUH

Propensity Stratum	Sample Size	Weighted Population Estimate before Adjustment (in Thousands)	Weighted Population Estimate after Adjustment (in Thousands)	Ratio of after to before Weighted Population Estimates
Hispanics Responding in English, by Modeled Propensity to Respond in English	19,745	88,197	127,311	1.4435
Very High (Upper 60%)	11,847	39,114	39,150	1.0009
High (Next 20%)	3,943	19,531	20,089	1.0286
Moderate (Next 10%)	1,979	13,005	17,607	1.3539
Low (Next 5%)	988	5,317	16,513	3.1057
Very Low (Lowest 5%)	988	5,569	33,953	6.0972
Hispanics Responding in Spanish	7,726	44,776	0	0.0000
All Hispanic Respondents	**27,471**	**127,311**	**127,311**	**1.0000**

NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2011.

Exhibit 5.1 Recruitment Screen for 2008-2012 MHSS Clinical Follow-Up

RECRUIT1

[IF ELIGIBLE FOR MENTAL HEALTH FOLLOW-UP STUDY] You have been randomly selected to participate in one additional study for the U. S. Public Health Service. This interview will ask questions about mental health issues. It will be conducted over the telephone and will take about an hour. Participation in this interview is voluntary and all of your answers will be confidential.

If you agree to participate, I will pay you an additional $30 today. Within the next two weeks, a different interviewer will call you to explain more about the interview and to schedule a convenient time to complete it. If you wish, you may complete the full interview when the interviewer calls.

HAND FOLLOW-UP STUDY DESCRIPTION TO RESPONDENT. Please read this statement. It describes the survey and the legislation that assures the confidentiality of any information you provide.

1 RESPONDENT AGREES TO RECONTACT
2 RESPONDENT DOES NOT AGREE TO RECONTACT
3 RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD.

RECRT4WK

[IF RECRUIT1=3] To accommodate your schedule, an interviewer will be available to call you about this study and schedule a convenient time to complete the interview within the next four weeks.

1 RESPONDENT AGREES TO RECONTACT
2 RESPONDENT DOES NOT AGREE TO RECONTACT
3 RESPONDENT IS NOT AVAILABLE DURING THE SPECIFIED TIME PERIOD.

Second, nonresponse occurs when the clinical interviewer attempts to contact the respondent who agreed to participate at the end of the main NSDUH interview. If there are very few differences between the nonrespondents in these two phases of participation in the MHSS clinical interview, they can be treated as a homogenous group in nonresponse adjustments. However, if there are differences between these groups of nonrespondents, adjustments tailored to each phase of nonresponse may have greater potential for reducing nonresponse bias.

Further distinctions among nonrespondents in the MHSS clinical study are possible but were not considered for evaluating differences between respondents and nonrespondents or for use in nonresponse weighting adjustments due to small sample sizes. Table 5.4 shows the distribution of result codes for nonrespondents for the clinical follow-up for 2008 to 2011. These are adults selected for the MHSS clinical follow-up who agreed to participate when asked at the end of the main NSDUH interview but then did not complete the clinical interview.[32]

Table 5.4 Result Codes for Adults Selected for Clinical Follow-Up Interviews and Agreeing to Participate, 2008 to 2011

Result Code	Number of Adults Selected	Percentage
Code 73 (Breakoff, Partial Interview)	157	14.2
Code 74 (Unable to Contact)	743	67.5
Code 75 (Phone Number Problem)	129	11.7
Code 76 (Refusal)	27	2.5
Code 77 (Other)	44	4.0
Total Incompletes	**1,100**	**99.9**

NSDUH = National Survey on Drug Use and Health.

NOTE: Due to rounding, the percentages do not add to 100 percent.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2011.

Most of the nonresponse at the point of attempting to conduct the clinical interview (67.5 percent) was classified as unable to contact. However, this classification does not mean that in all of these cases contact was not made with the selected adult. Many of these may have been passive refusals in which respondents did not actually refuse to participate but may have simply avoided subsequent contacts with the clinical interviewer or were not able to complete the interview within the 4 weeks allowed for completing clinical interviews. More important, only 2.5 percent of the nonresponse at this stage consisted of refusals. Thus, the causes of nonresponse at this stage can be different from that at the previous stage at the end of the main NSDUH interview, in which nonrespondents immediately refused to participate.

[32] Table 5.4 also includes clinical interview cases where the record was determined to be falsified (discussed in Section 3.8.4 in Chapter 3) and cases in 2008 through 2011 that should have been selected for the clinical interview based on their characteristics but were assigned a zero probability of selection instead. For more details, see Section 3.8.1 in Chapter 3.

5.4.2 Descriptive Characteristics of Adults Agreeing and Not Agreeing to Participate in the Clinical Follow-Up

Tables D2a to D2c in Appendix D present comparisons between NSDUH respondents agreeing to take part in the clinical follow-up and those who refused to do so. The weight used for these comparisons was the full base weight MHDSNWT for the clinical sample, which was the weight in the previous step that accounted for undercoverage of Hispanics interviewed in Spanish (ANALWT_E), adjusted for clinical sample selection probabilities.

Descriptive statistics comparing adults who agreed to participate in the clinical interview with initial nonrespondents are available from Table D2a:

- Those who agreed to participate in the clinical interview were generally younger than those who declined to participate. For example, 18.4 percent of those agreeing to participate were between the ages of 18 and 25, while only 8.6 percent of those refusing to participate were in that age range. Also, 51.6 percent of those refusing to participate were aged 50 or older, while only 38.3 percent of those who agreed to participate were aged 50 or older.
- Blacks made up 10.2 percent of those agreeing to participate, while blacks only made up 4.6 percent of those who refused to participate.
- An estimated 26.1 percent of adults who agreed to participate reported some college as their highest level of education as compared with 17.9 percent of adults refusing to participate in the clinical follow-up. There were no other statistically significant differences in the other levels of educational attainment.
- In terms of employment status, those refusing to participate were much more likely to be retired than those agreeing to participate (22.3 vs. 10.7 percent). Also, adults agreeing to participate were more likely to not be currently working but to have worked in the past 12 months than those refusing to participate (8.1 vs. 3.8 percent).
- Those refusing to participate were less likely to report ever having been diagnosed with depression (5.9 percent) than those agreeing to participate (15.1 percent). Also, those refusing to participate were more likely to report ever having been diagnosed with ulcers than those agreeing to participate (5.6 vs. 0.9 percent).
- Adults refusing to participate had shorter interview times than those agreeing to participate. For example, 49.9 percent of those refusing to participate completed the main NSDUH interview in less than 60 minutes as compared with only 27.7 percent of those who agreed to participate. Also, 77.2 percent of those agreeing to participate completed the adult depression module in less than 2 minutes, while 92.5 percent of those refusing to participate completed the depression module in less than 2 minutes.

Mental health measures are presented in Table D2b for comparisons of adults who agreed to participate in the clinical interview with initial nonrespondents:

- Consistent with the differences in the time needed to complete the adult depression module, 7.4 percent of those agreeing to participate reported MDE in the past year, while only 3.1 percent of refusers reported past year MDE.

- Also, 15.2 percent of those agreeing to participate reported receiving any mental health treatment in the past year, while only 8.7 percent of those refusing to participate reported receiving such treatment.

Substance use measures are presented in Table D2c for comparisons of adults who agreed to participate in the clinical interview with initial nonrespondents:

- Those refusing to participate had lower rates of lifetime and past year illicit drug use compared with those who agreed to participate in the clinical follow-up.
- For example, among the lifetime measures of illicit drug use, all differences between adults who refused to participate and those agreeing to participate were statistically significant, except for the nonmedical use of tranquilizers.

5.4.3 Development of the Adjustment for Adults Not Agreeing to Participate in the Clinical Follow-Up

Before adjusting for nonresponse, a few large values in the (coverage-adjusted) full base weights[33] (MHDSNWT) were trimmed so that (ideally) no single respondent in the clinical sample had more than a 4 percent impact on the direct estimates based on the clinical sample in a year and thereby undermined an assumption of large-sample probability-sampling theory.

A full base weight F_k greater than 4 percent of the sum of NSDUH analysis weights in the year was trimmed back to that value (T). This meant that the full base weight was multiplied by a trimming factor (MHADJ_2a):

$$A_{2ak} = min(1, T/F_k),$$

where A_{2ak} was defined to be 0 for NSDUH respondents not selected for the clinical interview.

The larger the yearly sample sizes and the less oversampling, the fewer the weights needing trimming. In total, nine records had their full base weights trimmed. Five were from 2009, two from 2008B, and one each from 2008A and 2010. The relatively large samples in 2011 and 2012 had none.

Selection for the follow-up clinical interview was done at the same time as the main NSDUH interview. A fraction of those selected for the clinical interview did not agree to be recontacted by telephone for a clinical interview. The initial refusal adjustment factor (MHADJ_2b) had a form very similar to equation (5.1):

$$A_{2bk} = \frac{1+\exp(\mathbf{g}^T\mathbf{z}_k)}{1+\exp(\mathbf{g}^T\mathbf{z}_k)/U}. \tag{5.3}$$

[33] These base weights are the weights of the full clinical sample before nonresponse adjustment with the coverage adjustment for the clinical sample as well as all the weighting adjustments in the NSDUH main survey person-level analysis weights.

The U in equation (5.3) capped the size of the adjustment and thus the impact a single record could have on estimation. It was set at 5 for most years and at 5.5 for 2009. That is as low as U could be set for the calibration equation (5.4), given in the following text, to be satisfied.

The components of $\mathbf{z}_k$ were selected from a much larger set of candidate variables. For a variable from the main NSDUH interview to be a candidate, there had to be a significant difference between its mean computed in the clinical sample selected to participate in the clinical follow-up using combined 2008 through 2011 clinical data and its mean in the subset of adults who agreed to be recontacted.

Starting with these candidate variables and using the same combined 2008-2011 clinical data, backward selection was applied to a logistic-regression analysis predicting whether a NSDUH respondent selected for the clinical interview would agree to be recontacted. The variables ultimately selected to be components of $\mathbf{z}_k$ in equation (5.3) had p values of less than 0.1. These components included the K6 score (0 to 24). The other components were (0/1) indicators for the following characteristics:

- age less than or equal to 25 years,
- black,
- at least some college,
- retired,
- worked in the past 12 months,
- lifetime depression,
- lifetime ulcer or ulcers,
- NSDUH main interview length of less than 60 minutes,
- NSDUH adult depression module of less than 2 minutes,
- past year MDE,
- lifetime cocaine or crack or heroin use,
- lifetime lysergic acid diethylamide (LSD) use,
- lifetime Ecstasy use,
- lifetime nonmedical use of psychotherapeutics,
- lifetime nonmedical use of pain relievers,
- lifetime methamphetamine use, and
- received any mental health treatment in the past year.

The vector $\mathbf{g}$ was chosen using the WTADJUST procedure to satisfy the calibration equation:

$$\sum_S F_k \mathbf{z}_k = \sum_S F_k A_{2ak} A_{2bk} \mathbf{z}_k, \tag{5.4}$$

where S was again a yearly NSDUH adult respondent sample and, by definition, $A_{2bk} = 0$ for a NSDUH respondent who was either not selected for the clinical interview or who refused to be recontacted when selected. This vector consistently estimates $\boldsymbol{\gamma}$ in the following model for the probability that an adult in the clinical sample agrees to be recontacted:

$$\rho_k = \frac{1 + \exp(\boldsymbol{\gamma}^T \mathbf{z}_k) / U}{1 + \exp(\boldsymbol{\gamma}^T \mathbf{z}_k)},$$

where an adult k selected for the clinical follow-up sample agreed to be recontacted is $p_k = 1/A_{2bk}$ when $A_{2bk} > 0$.

Table 5.5 shows how the weighting to adjust for those adults who refused to be recontacted after being selected for the clinical sample related to the estimated probabilities of agreeing to be recontacted. The adults who agreed to be recontacted were divided into five propensity strata based on their estimated probabilities of agreeing to be recontacted: highest 25 percent, next 25 percent, next 25 percent, next 15 percent, and lowest 10 percent. The table displays the ratio of the weighted population totals for adults after the weighting adjustment to before the weighting adjustment in each of the propensity stratum. The ratio is close to 1 for 75 percent of the selected sample (i.e., for the three highest propensity strata, divide the sum of 1,317, 1,318, and 1,317 by 5,271) and is as high as about 2 for the lowest response propensity stratum. This means that the weights for those adults in the lowest propensity stratum increased roughly 200 percent on average due to the weighting adjustment, while they hardly increased at

Table 5.5 Summary of Nonresponse Adjustments for Agreement to Participate in the Clinical Follow-Up, by Propensity Strata: 2008 to 2011 MHSS

Propensity Stratum	Sample Size	Weighted Population Estimate before Adjustment (in Thousands)	Weighted Population Estimate after Adjustment (in Thousands)	Ratio of after to before Weighted Population Estimates
Adults Agreeing, by Propensity to Agree to Participate in Clinical Follow-Up	**5,271**	**720,273**	**930,942**	**1.2925**
Very High (Upper 25%)	1,317	115,794	115,906	1.0010
High (Next 25%)	1,318	104,325	107,478	1.0302
Moderate (Next 25%)	1,317	182,829	200,206	1.0950
Low (Next 15%)	791	173,321	222,686	1.2848
Very Low (Lowest 10%)	528	144,003	284,666	1.9768
Sample Adults Not Agreeing to Participate in Clinical Follow-Up	894	210,669	0	0.0000
Total Sample	**6,165**	**930,942**	**930,942**	**1.0000**

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2011.

all for 75 percent of the adults in the three highest propensity strata. Those in the lowest propensity stratum (least likely to agree to participate in the clinical interview) had characteristics that were more similar to those who refused to participate than those in the higher propensity strata. Thus, adjusting their weights compensates for those who refused to participate in the clinical interview.

5.4.4 Descriptive Characteristics of Adults Completing and Not Completing the Clinical Follow-Up

Tables D3a to D3c in Appendix D present comparisons between respondents who completed the clinical follow-up and those who had agreed to participate in the clinical follow-up but did not. The weight used for these comparisons was MHWTNR1, which was the weight in the previous step that accounted for nonresponse by those who refused to take part in the clinical follow-up at the conclusion of the main NSDUH interview. Below are some highlights of these comparisons, particularly with respect to measures used in the adjustment for those not completing the clinical follow-up and items in the 2012 revised model for predicting SMI.

Descriptive statistics comparing adults who initially agreed to participate in the clinical interview with final nonrespondents are available from Table D3a:

- Women were more likely to complete the clinical follow-up interview than men. For example, 53.6 percent of those who completed the clinical interview were women, while men made up 60.6 percent of those who did not complete the clinical interview.
- Those who did not complete the interview were younger than those who did. Adults between the ages of 18 and 30 years old made up 23.0 percent of those who completed the clinical interview, but 36.6 percent of those who did not complete the clinical interview.
- Those who did not complete the clinical follow-up interview were less likely to be white and more likely to be Hispanic than those who completed it. White respondents made up 73.3 percent of those who completed the interview, but only 55.2 percent of those who did not complete it. Also, 12.4 percent of those who completed the interview were Hispanic, while Hispanics made up 27.8 percent of those who did not complete it.
- Respondents to the clinical follow-up interview reported higher levels of education than nonrespondents. For example, 29.2 percent of those who completed the interview were college graduates, while 16.1 percent of those who did not complete it were college graduates. Similarly, only 9.4 percent of those who completed the interview had less than a high school education, while 31.1 percent of those who did not complete it had less than a high school education.
- An estimated 30.4 percent of those who did not complete the clinical follow-up interview were not covered by health insurance. In contrast, only 14.0 percent of those who did complete it did not have health insurance.

- Respondents to the clinical follow-up interview were more likely to report no difficulties in understanding the main NSDUH interview (93.7 percent) compared with those who did not complete it (82.0 percent).

Mental health measures are presented in Table D3b for comparisons of adults who agreed to participate in the clinical interview with final nonrespondents:

- This table contains many of the items used in the 2012 revised model for predicting SMI. It is noteworthy that for measures such as K6 scores, WHODAS scores, past year MDE, and serious thoughts about suicide in the past year, there were no statistically significant differences between those who completed the clinical interview and those who did not.

Substance use measures are presented in Table D3c for comparisons of adults who agreed to participate in the clinical interview with final nonrespondents:

- Those refusing to participate had lower rates of lifetime and past year illicit drug use compared with those who agreed to participate in the clinical follow-up interview.
- For example, among the lifetime measures of illicit drug use, all of the differences between adults who refused to participate and those who agreed to participate were statistically significant except for the nonmedical use of tranquilizers.

Thus far, differences between respondents and nonrespondents have been presented in this section and the previous one. Differences between the two types of nonrespondents also are worth noting. Initial nonrespondents (i.e., those who refused to participate in the clinical follow-up interview immediately at the end of the main NSDUH interview) had lower prevalences for a number of items related to mental health compared with nonrespondents who initially agreed to participate in the clinical follow-up interview but ultimately did not complete that interview (i.e., the final nonrespondents). For example, 5.9 percent of initial nonrespondents reported ever having been diagnosed with depression (Table D2a), while 13.0 percent of final nonrespondents did so (Table D3a). The mean K6 score for initial nonrespondents was 3.2 (Table D2b), while it was 4.9 for final nonrespondents (Table D3b). Past year and lifetime rates of MDE were lower for initial nonrespondents than for final nonrespondents (Tables D2b and D3b).

There were also demographic differences between the two types of nonrespondents. Initial nonrespondents were more likely to be female than final nonrespondents (52.1 vs. 39.4 percent; Tables D2a and D3a). Initial nonrespondents were older than final nonrespondents. Only 13.5 percent of initial nonrespondents were aged 18 to 30 years old, while 36.6 percent of final nonrespondents were in that age range. Initial nonrespondents reported higher levels of education than final nonrespondents; 27.5 percent of initial nonrespondents were college graduates, while 16.1 percent of final nonrespondents were college graduates.

5.4.5 Development of the Adjustment for Adults Not Completing the Clinical Follow-Up

The adjustments for nonresponse among adults who had originally agreed to be recontacted for clinical follow-up paralleled the refusal adjustments in the previous

adjustment step. First, seven records had their adjusted weights for initial refusal (MHWTNR1) trimmed to T (the trimming factor is MHADJ_3a). Three were from 2009, two from 2008B, and one each from 2008A and 2010.

The nonresponse adjustment factor (MHADJ_3b) for not completing the clinical interview employed an unbounded (by U) version of equation (5.1):

$$A_{3bk} = 1 + \exp(\mathbf{g}^T \mathbf{z}_k). \tag{5.5}$$

The components of $\mathbf{z}_k$ were selected for equation (5.5) from a large set of candidate variables (which were again chosen based on univariate comparisons) by fitting a logistic regression to adults who had agreed to be recontacted for the clinical interview from 2008 through 2011. The dependent variable for this model was whether or not a selected adult completed the clinical interview. Each component had a p value of less than 0.1.

The components of $\mathbf{z}_k$ were 0/1 indicators of the following characteristics:

- female,
- white,
- less than high school,
- covered by any health insurance,
- no difficulty understanding the main NSDUH interview,
- analgesic dependence or abuse,
- alcohol dependence or abuse,
- lifetime sedative use,
- past year cigarette use,
- NSDUH main interview of less than 60 minutes, and
- lifetime nonmedical use of psychotherapeutics.

The vector $\mathbf{g}$ was chosen using WTADJUST to satisfy the calibration equation:

$$\sum_S F_k A_{2ak} A_{2bk} \mathbf{z}_k = \sum_S F_k A_{2ak} A_{2bk} A_{3ak} A_{3bk} \mathbf{z}_k, \tag{5.6}$$

where S was again a yearly NSDUH adult respondent sample, A_{3ak} the trimming factor applied to the weight after initial refusal, if necessary, and $A_{3bk} = 0$ for any NSDUH respondent who was not a respondent in the clinical follow-up study. This vector consistently estimates $\boldsymbol{\gamma}$ in the following model for an adult in the clinical sample who agreed to be recontacted actually completing the clinical interview:

$$\rho_k = \frac{1}{1 + \exp(\boldsymbol{\gamma}^T \mathbf{z}_k)}.$$

The estimated probability that an adult k who agreed to be reinterviewed actually completed the clinical follow-up interview is $p_k = 1/A_{3bk}$ when $A_{3bk} > 0$.

Table 5.6 shows how the weighting to adjust for adults who agreed to be recontacted but failed to complete the clinical follow-up reinterview related to the probabilities of completing the clinical interview among those who agreed to be recontacted. The adults who completed the clinical interview were divided into five propensity strata based on their estimated probabilities of completing the clinical interview: highest 25 percent, next 25 percent, next 25 percent, next 15 percent, and lowest 10 percent. The table displays the ratio of the weighted population totals for adults after the weighting adjustment to before the weighting adjustment in each of the propensity strata. The ratio is close to 1 for 75 percent of the adults who completed the clinical follow-up interview (i.e., for the highest three propensity strata). By contrast, for the lowest propensity stratum, it is close to 3. This means that the weights for those adults in the lowest response propensity stratum increased roughly 300 percent on average because of the weighting adjustment, while they increased hardly at all for 75 percent adults in the highest three propensity strata. Adults in the lowest propensity stratum (i.e., those least likely to complete the clinical follow-up given that they initially agreed to participate) had characteristics most similar to adults who initially agreed but failed to complete the clinical follow-up. Thus, increasing the weights of adults in the lowest propensity stratum compensates for those clinical-interview nonrespondents.

Table 5.6 Summary of Nonresponse Adjustments for Adults Completing the Clinical Follow-Up Given That They Agreed to be Recontacted, by Propensity Strata: 2008 to 2011 MHSS

Propensity Stratum	Sample Size	Weighted Population Estimate Before Adjustment (in Thousands)	Weighted Population Estimate After Adjustment (in Thousands)	Ratio of After to Before Weighted Population Estimates
Adults Completing Clinical Follow-Up, by Propensity to Complete Clinical Follow-Up	**4,031**	**715,984**	**930,942**	**1.3002**
Very High (Upper 25%)	980	164,197	168,978	1.0291
High (Next 25%)	1,034	193,532	215,857	1.1154
Moderate (Next 25%)	1,009	206,169	252,779	1.2261
Low (Next 15%)	604	107,795	163,737	1.5190
Very Low (Lowest 10%)	404	44,291	129,591	2.9259
Sample Adults Initially Agreeing but Not Completing Clinical Follow-Up	1,240	214,958	0	0.0000
Total Sample	**5,271**	**930,942**	**930,942**	**1.0000**

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2011.

5.4.6 Summary of Nonresponse Adjustments

It is worth noting that there is very little overlap between the two sets of measures used to adjust for the two different phases of nonresponse. The measures that are predictive of agreement to participate in the clinical interview at the end of the main NSDUH interview are mostly different from the items predictive of completing the clinical follow-up interview, conditional on having agreed to participate. The only items in common between the two sets of variables used for nonresponse adjustments are race/ethnicity, education, NSDUH main interview length, and lifetime nonmedical use of psychotherapeutics. The adjustment for adults refusing to participate at the end of the main NSDUH interview contains 13 other variables, and the adjustment for not completing the clinical interview contains 7 other variables. This suggests that the causes of these two types of nonresponse are distinct from each other and that the adjustments should be carried out separately.

5.5 Development of Revised Poststratification Adjustment and Scaling

5.5.1 Poststratification

The term "poststratification" has been used to describe a weighting step that forces the weighted totals of a set of variables to equal the population totals provided from an outside source such as the U.S. Census Bureau.[34] For the NSDUH main survey and previously for the clinical sample under the 2008 MHSS modeling process (described in Liao et al., 2012, Chapter 4), such a poststratification step was used to adjust for undercoverage in the sampling frame. In developing the new clinical sample weights, however, it was assumed that both coverage errors and nonresponse biases were removed prior to conducting the poststratification step. The goal of this step was to decrease the SEs of estimates for mental health characteristics estimated directly from the clinical sample. This was done by forcing weighted totals after adjustment to equal totals not provided from an outside source per se but estimated from the overall NSDUH main survey adult respondents.

The poststratification process began with the adjusted weight for a completed clinical interview (MHWTNR2), $q_k = F_k A_{2ak} A_{2bk} A_{3ak} A_{3bk}$. First, six records had their weights trimmed to T as in the previous two steps. Three were from 2009 and one each from 2008A, 2008B, and 2010.

In previous adjustment steps, the adjustment factor implicitly estimated the probability of responding to the NSDUH main interview in English or the probability of responding to the clinical follow-up interview. In this poststratification step, no such probability is being estimated. In fact, the adjustment factor can be viewed as an "estimate" of 1. The poststratification weighting adjustment had the following form:

[34] See Chen et al. (2014) for a description of the poststratification step in developing the NSDUH main survey person-level analysis weights. This use of the term "poststratification" does not strictly conform to the textbook definition in, for example, Lohr (1999, pp. 113-114), which requires target population totals to be for mutually exclusive groups (and so limits the number of types of groups that could be used in the step).

$$A_{4bk} = \frac{\exp\left\{[(U_k - 1)/U_k]a_k \mathbf{g}^T \mathbf{z}_k\right\}}{(U_k - 1)/U_k + \exp\left\{[(U_k - 1)/U_k]a_k \mathbf{g}^T \mathbf{z}_k\right\}\Big/U_k}, \tag{5.7}$$

where $a_k = q_k/w_k$ and $U_k = T/q_k$ in all years but 2008A, where it was relaxed to 1.25 T/q_k for the calibration equation (below) to hold.[35] This setting bounded the fully adjusted weights themselves (MHNEWWGT) to T where possible and 1.25 T otherwise. Unlike in the previous adjustments, the choice of A_{4bk} was not related to the probability that adult k responded to the NSDUH main survey in English or successfully completed the clinical sample. As a result, the bounding parameters U_k were allowed to vary from one adult to another.

In equation (5.7), the vector $\mathbf{z}_k$ consists of the variables chosen because their inclusion in $\mathbf{z}_k$ decreased the SEs of estimates for past year SMI, any mental illness (AMI), and MDE computed from the clinical sample without the aid of a model. These variables are as follows:

- indicators for six categories of gender (male and female) by age (18 to 25, 26 to 34, 35 or older) categories,
- indicators for race/ethnicity categories (Hispanic, non-Hispanic white, non-Hispanic black, other),
- an indicator for suicidal thoughts,
- indicators from the NSDUH main interview for a past year and lifetime MDE,
- the interaction term between an alternative K6 score (defined as the maximum value between 0 and K6SCMAX – 7)[36] and three age categories, and
- the interaction term between an alternative WHODAS score (ranges from 0 to 8) and three age categories.[37]

The vector **g** in equation (5.7) was found using the WTADJX[38] procedure in SUDAAN 11.0 (RTI International, 2012) to satisfy the calibration equation:

---

[35] Equation (5.7) follows the bounded pseudo-optimal calibration weighting discussed in Kott (2011).

[36] See Chapter 2 for a discussion of the alternative K6 score.

[37] See Chapter 2 for a discussion of the alternative WHODAS score. For the 2008B sample, it was replaced by an alternative SDS measure.

[38] The insertion of the a_k into equation (5.7), suggested in Kott (2011), tends to decrease the SEs of the resulting estimators (they would still decrease without the a_k, but not by as much). Because of it, WTADJUST could no longer be used (nor could the GEM macro) and was replaced by WTADJX. Implementing WTADJX with ADJUST=POST in this context requires some work. First, a dataset is created with both main NSDUH and clinical sample respondents. Adults in the latter group are listed twice in this dataset, once as members of the main sample and once as members of the clinical sample. A main sample member has an initial weight of w_k, while a clinical sample member has an initial weight of $\eta_{kk} = F_k A_{2ak} A_{2bk} A_{3ak} A_{3bk} A_{4ak}$. The components of $c_k\mathbf{z}_k$ are in the CALVARS statement, where $c_k = 1$ for the clinical sample members, and $c_k = -1$ for the main sample members. The MODEL statement is _ONE_ = the components of $m_k\mathbf{z}_k$, where $m_k = a_k$ for the clinical sample members, and $m_k = 0$ for the main sample members. The components of POSTWGT are all 0. The lower bound is 0, and the upper bound is U_k.

$$\sum_S w_k \mathbf{z}_k = \sum_S F_k A_{2ak} A_{2bk} A_{3ak} A_{3bk} A_{4ak} A_{4bk} \mathbf{z}_k, \tag{5.8}$$

where A_{4ak} was the trimming factor applied to the weight after other nonresponse adjustments, if necessary, and $A_{4bk} = 0$ for any NSDUH respondent who was not a respondent in the clinical follow-up study.

Table 5.7 presents the poststratification controls used in the adjustment process along with the ratio of the summed weights before and after adjustment. The ratio can be viewed as the relative degree to which weights were increased or decreased from the previous weighting step to meet the desired control totals as specified in the poststratification. The groups shown are those used in the poststratification. For targets that involved noncategorical measures (e.g., the K6, WHODAS, or SDS scores), means are provided with the age categories that were used in the adjustment as well as the ratios between these means before and after adjustment. For these, a ratio greater than 1 indicates that adults with higher K6, WHODAS, or SDS scores had their weights increased by the poststratification, while a ratio of less than 1 indicates that adults with higher scores on these measures had their weights decreased. For most groups, the ratios of the weighted population totals from after to before poststratification are fairly close to 1, indicating that on average, the cases in these groups did not have their weights increased or decreased by much due to the poststratification.

5.5.2 Scaling

Because of the variations in the clinical interview sample sizes, as well as the sample allocations and weighting adjustments, a scaling factor was applied to weights in each year before combining the years. This was applied to reduce the variance of the estimates for mental health characteristics estimated directly from the clinical diagnoses and to increase the efficiency in the modeling described in Chapter 7. The scaling factors were determined based on the reduction of the variance in the estimates of SMI, AMI, and past year MDE that were made directly from SCID diagnoses. For the combined 2008-2012 clinical sample including both the 2008A and 2008B clinical samples, analysis weights, designated as MHFNLWGT, were created by scaling the MHNEWWGT weights in each year with the following scaling factors: 6 percent for 2008A sample, 6 percent for 2008B sample, 4 percent for 2009, 14 percent for 2010, 35 percent for 2011, and 35 percent for 2012.

For the combined 2008-2012 clinical sample including only the 2008A sample, the weight MHNEWWGT was scaled by using the following scaling factors: 12 percent for the 2008A sample, 4 percent for 2009, 14 percent for 2010, 35 percent for 2011, and 35 percent for 2012. This set of scaled weights, designated as MHFAAWGT, was applied in the development of the 2012 model.

Table 5.7 Poststratification Controls and Average Adjustment: 2008 to 2011 MHSS Clinical Sample

Description	Value(s)	Weighted Population Estimate before Adjustment (in Thousands)	Weighted Population Estimate after Adjustment (in Thousands)	Ratio of after to before Weighted Population Estimates
Gender	Male	462,212	441,105	0.9543
	Female	468,731	472,923	1.0089
Age	18 to 25	163,236	134,892	0.8264
	26 to 49	381,976	395,813	1.0362
	50 or Older	385,731	383,322	0.9938
Gender, by Age	Male, 18 to 25	89,423	67,916	0.7595
	Male, 26 to 49	200,597	195,318	0.9737
	Male, 50 or Older	172,191	177,871	1.0330
	Female, 18 to 25	73,813	66,977	0.9074
	Female, 26 to 49	181,379	200,494	1.1054
	Female, 50 or Older	213,539	205,451	0.9621
Race	White	643,530	621,345	0.9655
	Black	99,531	105,144	1.0564
	Hispanic	126,260	127,311	1.0083
	Asian and Others	61,622	60,227	0.9774
Seriously Think about Killing Self in Past 12 Months	Yes	36,226	34,079	0.9407
Lifetime MDE	Yes	125,099	117,823	0.9418
Past Year MDE	Yes	65,985	60,811	0.9216
Overall		930,942	914,027	0.9818
Alternative K6 Score, by Age[1]	18 to 25	2.20	2.35	1.0711
	26 to 49	1.77	1.62	0.9167
	50 or Older	0.94	0.88	0.9358
Alternative WHODAS Score, by Age[2]	18 to 25	1.27	1.18	0.9311
	26 to 49	1.04	0.93	0.8958
	50 or Older	0.55	0.66	1.2038
Alternative SDS Score, by Age[3]	18 to 25	0.29	0.40	1.3753
	26 to 49	0.26	0.34	1.2807
	50 or Older	0.24	0.19	0.7938

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SDS = Sheehan Disability Scale; WHODAS = World Health Organization Disability Assessment Schedule.

NOTE: The weighted population estimates before adjustment used WEIGHT = MHWTNR2; the weighted population estimates after adjustment used WEIGHT = MHNEWWT.

[1] Alternative K6 score is defined as max(0, K6SCMAX – 7).
[2] Alternative WHODAS total score (range 0–8), for 2008 sample A and 2009-2011.
[3] Alternative SDS total score (range 0–4), for 2008 sample B.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2011.

5.6 Evaluation of the Final Weights

5.6.1 Comparisons with the NSDUH Main Adult Sample

To evaluate how well the revised weighting adjustments have addressed bias in the clinical sample, estimates for NSDUH items based on different samples with different weights from 2008 through 2011 are presented in Tables E1 to E3 in Appendix E. The first data column in each of these tables displays estimates from the NSDUH main sample for adults using the NSDUH (main) analysis weights. Tests of statistical significance were carried out between estimates in each of the other data columns and estimates in the first data column. Thus, estimates from the full NSDUH sample are treated as the benchmark for judging the impact of the various weighting adjustments. There is no comparison with estimates based on scaled clinical sample weights because the NSDUH analysis weights are not scaled.

Columns 2, 3, and 4 in Tables E1 to E3 present estimates for adults selected for the clinical interview, nonrespondents (at both phases of nonresponse), and respondents, respectively. All of these estimates use the same weight, MHDSNWT (see Table 5.1 for a description of the weights), which includes the NSDUH adult-level analysis weights, the adjustment for undercoverage of Hispanics[39] who completed the main NSDUH interview in Spanish, and the probability of selection for the clinical interview. Tests of statistical significance were carried out for differences in estimates between nonrespondents (column 3) and respondents (column 4) prior to adjustment for nonresponse.

The final two columns of the tables present estimates for NSDUH items using the revised weights derived from the revised weighting adjustments before scaling (MHNEWWGT) (column 5) and the previous weights derived from the 2008 weighting adjustments (column 6). In addition to comparisons with the "benchmark" estimates from the full NSDUH sample in column 1, the estimates in columns 5 and 6 were also compared with each other to see whether the revised weighting produced any changes in estimates compared with the previous weighting adjustments.

Across all three tables, statistically significant differences between the respondents and nonrespondents in the clinical sample are evident for many items. The descriptive characteristics shown in Table E1 include such demographics as gender, age, race/ethnicity, education, income, region, employment, health insurance coverage, ever having been diagnosed with depression, and country of birth. The mental health items in Table E2 include measures related to K6 scores, WHODAS scores, past year and lifetime measures of MDE, and measures related to suicidal thoughts and mental health treatment. Some of these variables are used in the revised 2012 model for predicting SMI or, at least on the surface, may be correlated with specific disorders assessed on the SCID. The substance use measures shown in Table E3 provide fewer statistically significant differences between respondents and nonrespondents as compared with items shown in Tables E1 and E2.

[39] Tables in Appendix D address the impact of the removal of adults responding to the NSDUH main survey in Spanish from the clinical survey's target population and the effectiveness of weighting to compensate for that.

In addition, the three tables allow comparisons of estimates to be made based on the previous and revised weights. For the largely demographic items in Table E1 and the substance use measures in Table E3, relatively few statistically significant differences can be seen between estimates using the previous and revised weights. Several statistically significant differences are worth noting, however, in the mental health items shown in Table E2. The estimate for lifetime MDE was lower using the revised weights than using the previous weights (12.9 vs. 14.6 percent). Furthermore, the revised weights' MDE estimate was close to the full sample NSDUH estimate (12.8 percent) and was not significantly different. The difference between the estimate using the previous weights (14.6 percent) and the full sample NSDUH estimate, however, was statistically significant. The mean K6 score using the revised weights was 5.1. This was not different (statistically) from the NSDUH full sample estimate of 4.9. However, the mean K6 score of 5.3 under the previous weights was significantly different from the full sample estimate. It should be noted that K6 scores (alternative) were added to the poststratification step for the revised weights.

There appears to be a tendency for the estimates of the respondents prior to adjustment (Table E2, column 4) to more closely resemble the estimates for the final nonrespondents in Table D3b than the initial nonrespondents in Table D2b. For example, the mean K6 score for respondents prior to adjustment was 5.2, while it was 4.9 for final nonrespondents but only 3.2 for initial nonrespondents. Similarly, for lifetime MDE, the estimates were 14.7 percent for respondents prior to adjustment, 12.7 percent for final nonrespondents, and only 5.4 percent for initial nonrespondents.

For a number of measures on mental health treatment and MDE treatment, estimates with the revised weights were to some extent closer to the estimates from the full NSDUH sample than those with the previous weights. These measures were not used in the poststratification adjustment, and most were not used in any of the undercoverage or nonresponse adjustment processes. For example, 6.6 percent of adults aged 18 or older received outpatient mental health treatment in the past 12 months in the full NSDUH sample (Table E2). The estimate for respondents only was 7.7 percent. The revised weights yielded an estimate of 6.9 percent, which was not significantly different from the full NSDUH sample's estimate. In contrast, the previous weights produced an estimate of 7.9 percent, which was significantly different from the estimate for the full NSDUH sample. Other measures in which the estimate produced by the revised weights appeared closer to the full NSDUH sample's estimate than the estimate from the previous weights include receiving any mental health treatment in the past year, using prescription drugs for MDE in the past year, and an unmet need for mental health treatment in the past year.

5.6.2 Unequal Weighting Factors

Another factor to consider in evaluating the revised weighting adjustments is the degree to which the variances of estimates may be affected by the variability of the weights themselves. Of interest here is variation introduced by the implicit use of (sometimes truncated) logistic regression modeling to derive weighting adjustments. When using a modeling approach, adjustment factors may be unstable, which can lead to wide variability in the weights. The unequal weighting effects (UWEs), which are an indicator of the variability of the weights,

are presented in Table 5.8 for each of the weight components examined here. The UWE for year y is as follows:

$$UWE_y = \frac{n_y \sum_{i \in y} w_i^2}{\left(\sum_{i \in y} w_i\right)^2},$$

where w_i represents the weights derived from the revised weighting adjustments before scaling (MHNEWWGT) and n_y is the number of clinical interview respondents in year y.

As a point of comparison, the UWE for the final adult-level weight (ANALWT, or equivalently ANALWT_A) in the 2011 NSDUH for all adults was 2.87. Prior to any adult-level adjustments for the NSDUH main sample (including nonresponse adjustment and poststratification), the UWE was 2.53 (Chen et al., 2014), which implies that the adult-level adjustments for the NSDUH main study inflated the variability of the NSDUH weights. Similarly, the UWE for ANALWT_A increased after the undercoverage adjustment for Hispanics who were interviewed in Spanish (resulted in ANALWT_E) for each year of NSDUH from 2008 to 2012.

Based on the clinical sample, the UWEs for ANALWT_A and ANALWT_E were similar to their counterparts based on the NSDUH main sample, but increased after incorporating the clinical sample selection probability that resulted in the full base weight of the clinical sample (MHDSNWGT). This increase was due to the unequal probability sample design for the clinical sample. Moreover, the UWE for MHDSNWGT generally increased after the first nonresponse adjustment (resulted in MHWTNR1), except for 2012 in which the UWEs for MHDSNWGT and MHWTNR1 were similar (12.04 vs. 11.55). The UWEs for the MHSS clinical sample generally declined from MHWTNR1 to the final analysis weight prior to rescaling (MHNEWWGT). One exception occurred in 2012 when the UWE for the second nonresponse adjustment (MHWTNR2) was a little higher than the UWE for MHWTNR1 (3.74 vs. 3.67). The results suggest that the second nonresponse adjustment and final poststratification were able to smooth the first nonresponse adjustment weights and reduce the weight variability. Also, some of the changes over time in the UWE for each weight component can be accounted for by changes in the sample allocation. For example, the UWE sharply decreased between the 2009 and 2010 surveys, which resulted from a change in the sampling algorithm in which adults in older age groups were oversampled and those in younger age groups were undersampled, which roughly reversed the oversampling in the NSDUH main survey.

5.6.3 Comparisons Based on Standard Error Computations

This section looks at the cumulative impact of the weight adjustments on the SEs of direct estimates computed from the clinical subsample. This analysis incorporates not only the impact on SEs resulting from the UWEs, but also from the poststratification. The former tends to increase SE, while the latter decreases SE. Because the focus is only on the SEs, whatever impact the weight adjustments may have on bias is not considered in the analysis.

Table 5.8 Unequal Weighting Effects for Weights at Stages of Adjustment, by Year of Data Collection: 2008 to 2012 Adult NSDUH Main Study and Clinical Sample

	Main Study		Clinical Sample					
Year/Sample	NSDUH Adult-Level Analysis Weight (ANALWT_A)	Undercoverage Adjustment for Hispanics Interviewed in Spanish (ANALWT_E)	NSDUH Adult-Level Analysis Weight (ANALWT_A)	Undercoverage Adjustment for Hispanics Interviewed in Spanish (ANALWT_E)	Full Base Weight for Clinical Sample (MHDSNWT)	Nonresponse Adjustment for Initial Nonresponse (MHWTNR1)	Nonresponse Adjustment for Final Nonresponse (MHWTNR2)	Analysis Weight for Clinical Sample, Prior to Rescaling (MHNEWWGT)
2008A Count	22,622	21,598	1,194	1,194	1,194	1,024	759	759
2008A UWE	3.15	3.61	3.17	3.28	10.64	13.00	11.11	10.96
2008B Count	23,046	21,991	1,137	1,137	1,137	996	741	741
2008B UWE	3.00	3.92	3.16	3.57	9.94	11.78	9.45	10.46
2009 Count	45,609	43,710	789	789	789	686	520	520
2009 UWE	2.92	3.45	2.73	4.48	12.04	11.55	8.24	7.46
2010 Count	45,844	43,961	768	768	768	644	516	516
2010 UWE	2.98	3.80	2.30	3.39	4.78	5.24	3.94	3.36
2011 Count	46,599	44,734	2,277	2,277	2,277	1,921	1,495	1,495
2011 UWE	2.87	3.52	2.29	3.12	4.06	4.12	3.73	3.07
2012 Count	45,836	44,228	2,464	2,462[1]	2,462[1]	2,062	1,622	1,622
2012 UWE	2.88	3.41	2.11	2.69	3.39	3.67	3.74	3.44

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; UWE = unequal weighting effect.

[1] Two Hispanics selected for the MHSS clinical interview in 2012 had responded to part of the NSDUH main interview in Spanish. As a result, they were treated as if they had been ineligible for the clinical interview when weighting the data (i.e., they were assigned ANALWT_E values of 0).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Using the scaled weights described in Section 5.5.2, Table 5.9 displays direct estimates from the clinical sample for SMI, AMI, and past year MDE using both the final weights (MHFNLWGT) and the "unadjusted" weights (ANALWT_A × MHWT1). The latter treat all sources of nonrandom missingness—whether due to nonresponse or being excluded from the target population—as if they could be ignored.

Table 5.9 Direct Estimates Based on SCID Sample: 2008-2012 MHSS

	Final Weights		Unadjusted Weights[1]		Difference	
Variable	**Mean**	**SE[2]**	**Mean**	**SE**	**Mean**	**SE**
SMI	3.94	0.24	4.13	0.26	-0.19	-0.02
AMI	17.95	0.66	18.36	0.81	-0.40	-0.14
MDE	6.32	0.36	7.12	0.56	-0.79	-0.20

AMI = any mental illness; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP); SE = standard error; SMI = serious mental illness.

NOTE: Estimates in this table were combined across years after scaling the weights as described later in the section.

[1] Estimated means and SEs computed using as the product of the NSDUH main study weight and the inverse of the clinical selection probabilities as the weight. This assumes that missingness whether due to undercoverage or nonresponse was at random.

[2] Computed by applying WTADJX to the combined 2008-2012 data, using scaled versions of w_k and η_k from footnote 38 and separate calibration (CALVARS) and model variables for each year (2008A and 2008B were treated as distinct years).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

One can see that the net impact of the weighting adjustment is that the estimates are lower. The table also computes SEs for both sets of estimates under the assumption that they are not subject to systematic biases (i.e., that the adjustments made to remove potential biases due to the undercoverage of Hispanics and clinical sample nonresponse were unnecessary; this assumption is only required for the unadjusted weights). The SE measures for the estimates computed using the adjusted weights properly account for the impact of the poststratification step,[40] but not fully for the other weighting adjustment steps. That is to say, it accounts for the impact of the steps on the UWE, but not for any further reduction of SE that forcing equations (5.2), (5.4), or (5.6) to hold might cause.

5.7 Summary

Weights are created and used in estimation in an attempt to remove the potential for systematic biases in sample survey estimates. Ideally, a survey respondent's weight is the inverse of his or her probability of selection into the sample, which is known to the statistician drawing the sample. In practice, this weight often needs to be adjusted to account for unit (whole-record) nonresponse to the survey and/or errors in the frame from which the sample was drawn.

[40] WTADJX does this by treating the weighted (by ω_k) mean of yk in the clinical sample as the weighted (by w_k) mean of $\mathbf{z}_k^T\mathbf{b} + (\omega_k/w_k)(y_k - \mathbf{z}_k^T\mathbf{b})$ in the NSDUH main sample, where $\omega_k = \eta_k A_{4bk} = F_k\, A_{2ak}\, A_{2bk}\, A_{3ak}\, A_{3bk}\, A_{4ak}\, A_{4bk}$ is the final clinical weight, $\mathbf{b} = (\sum_S \eta_k A_{4bk}' a_k \mathbf{z}_k \mathbf{z}_k^T)^{-1} \sum_S \eta_k A_{4bk}' a_k \mathbf{z}_k y_k$, and A_{4bk}' is the derivative of the right-hand side of equation (5.7) with respect to $a_k \mathbf{z}_k^T \mathbf{g}$. Kott (2011) pointed out that $\eta_k A_{4bk}'$ is asymptotically identical to 1 and η_k is asymptotically identical to ω_k.

The NSDUH main survey person-level weight incorporated adjustments as well as probabilities of selection. The clinical follow-up survey of adults also required additional adjustments: for the exclusion of adult NSDUH respondents who completed the main interview in Spanish, for those adults who were selected for the clinical follow-up interview but refused to be recontacted, and for those adults who agreed to be recontacted but failed to complete the clinical interview.

The clinical study weighting scheme made separate adjustment for each of these reasons and added a fourth adjustment to reduce the SEs of estimates derived from the clinical sample. A scaling factor was applied to the revised weights in each year before combining the data across years because the clinical interview sample sizes, as well as the sample allocations and weighting adjustments, varied from year to year. Applying this scaling reduced the variance of estimates for mental health characteristics estimated directly from the clinical sample and was intended to increase the efficiency of the modeling that is described in Chapter 7.

6. MHSS Clinical Sample Characteristics

6.1 Background

Both the annual weighted and unweighted Mental Health Surveillance Study (MHSS) clinical interview samples were evaluated each year from 2008 to 2012. The unweighted samples were evaluated to ensure that the sample composition reflected the sampling algorithm (e.g., the oversampling of higher Kessler-6 [K6] scores in 2008). The weighted samples were evaluated to ensure that the weighted distribution of the demographic and other characteristics of the clinical sample was similar to the National Survey on Drug Use and Health (NSDUH) main interview data.

This evaluation was done by comparing demographic and key mental health and substance use characteristics in the clinical sample across years. This chapter provides a summary of the single-year evaluations that were conducted and a description of the combined 2008 to 2012 clinical data.[41]

6.2 Summary of Prior Sample Evaluations

As described in Chapter 3, a split-sample design was implemented in 2008 for NSDUH in which a random half of the sample received an abbreviated version of the World Health Organization Disability Assessment Schedule (WHODAS) and the other half received the Sheehan Disability Scale (SDS). The randomization of the impairment scales was maintained within this clinical interview subsample, so that about half of the clinical interview participants were administered the WHODAS and half were administered the SDS. Therefore, the 2008 MHSS clinical sample evaluation consisted of descriptive analyses and statistical tests that were conducted to examine the distribution of respondent characteristics in the full clinical sample and to compare the two half samples, The purpose of these analyses was to determine whether estimates from the two half samples were comparable without accounting for differences caused by random sampling between the two samples. Key demographic characteristics examined included gender, age, race/ethnicity, and education. Substance use characteristics included past month tobacco and marijuana use and past year alcohol abuse or dependence. The mental health characteristics included Structured Clinical Interview for DSM-IV (SCID) diagnoses, serious mental illness (SMI) status, and K6 scores. Further details about these variables can be obtained from the 2008 NSDUH national findings report (Office of Applied Studies [OAS], 2009).

Analyzing both the unweighted and weighted data from the 2008 MHSS clinical study indicated that none of the demographic or mental health measures were significantly different across the two samples. There were, however, some large differences in the substance use measures, including past month tobacco and marijuana use. For details, see Aldworth et al. (2009).

[41] The weights used in the prior descriptive analyses were the old MHSS analysis weights (e.g., Liao et al., 2012), whereas the 2012 descriptive analyses used the new MHSS analysis weights developed in 2013. Chapter 5 provides more details about the components of these new weights.

For the 2009 MHSS clinical sample evaluation, initial descriptive analyses and statistical tests compared key demographic and psychosocial characteristics between the 2009 MHSS clinical sample and the 2008 MHSS WHODAS clinical sample. The purpose of these analyses was to determine whether the 2009 sample characteristics (see Section 3.5 for details) were consistent with the data collection from the WHODAS half sample of 2008. Key demographic characteristics included gender, age, race/ethnicity, and education. Mental health characteristics included lifetime and past year depression, depression treatment, mental health treatment, and suicidality measures. Substance use characteristics included past month tobacco and marijuana use and past year alcohol abuse or dependence. Additional mental health characteristics from the SCID (i.e., the MHSS clinical interview) included SMI status and diagnoses of mental health and substance use. Further details about these variables can be obtained from the 2009 NSDUH mental health findings report (Center for Behavioral Health Statistics and Quality [CBHSQ], 2010) and the 2009 NSDUH national findings report (OAS, 2010a, 2010b).

None of the unweighted demographic characteristics were significantly different between the two samples (i.e., the 2009 sample and the WHODAS half sample of 2008), but some of the prevalence estimates of mental health and substance use measures differed significantly. After applying the weights, most of these differences were no longer statistically significant. There were, however statistically significant (i.e., $p < 0.05$) and marginally significant (i.e., $p < 0.10$) differences, respectively, for past month cigarette use and past year alcohol abuse and dependence between the two samples. Differences in the past year K6 total scores were statistically significant before the weights were applied, but they were no longer significantly different once the data were weighted. For details, see Aldworth et al. (2010).

For the 2010 MHSS clinical sample evaluation, initial descriptive analyses and statistical tests compared key demographic and psychosocial characteristics in the clinical sample that were collected in 2008 and the first two quarters of 2009 with the clinical sample collected in the final two quarters of 2009 and the full clinical sample from 2010. The purpose of this particular comparison was to determine whether the expected changes based on the revisions in sampling allocation design after the second quarter of 2009 were being realized (see Sections 3.5 and 3.6 for details) and whether the weights were appropriately accounting for these changes. Further changes to the sampling design were made in 2010. To determine whether changes in the sampling design affected the demographic, mental health, and substance use characteristics between the samples, comparisons of these measures were made between the final two quarters of 2009 and 2010. Key demographic characteristics included gender, age, race/ethnicity, and education. Mental health characteristics included lifetime and past year depression, depression treatment, mental health treatment, and suicidality measures. Substance use characteristics included past month tobacco and marijuana use and past year alcohol abuse or dependence. Additional mental health characteristics from the SCID included SMI status and diagnoses of mental health and substance use. Details about these variables can be obtained from the 2010 NSDUH national findings report (CBHSQ, 2011) and the 2010 NSDUH mental health findings report (CBHSQ, 2012a).

Between the first two time periods and the final two time periods, unweighted estimates of age and education were significantly different. Some of the prevalence estimates of the mental health and substance use measures also appeared to be significantly different between the first two time periods and the final two time periods. Once the data were weighted, however, there

were fewer significant differences. Specifically, only the difference of past year alcohol and illicit drug abuse or dependence was still statistically significant ($p < 0.05$). The K6 scores within the SCID cases (i.e., the MHSS clinical interview cases) were compared between the two samples by Cochran-Mantel-Haenszel (CMH) tests of general association (24 *df* [degrees of freedom]) and means (1 *df*). For the unweighted scores, the two tests were statistically significant, but for the weighted scores only the test of general association was statistically significant. For details, see Aldworth et al. (2012).

For the 2011 MHSS clinical sample evaluation, initial descriptive analyses and statistical tests compared key demographic and psychosocial characteristics between the WHODAS sample of 2008-2010 with the WHODAS sample of 2011. The purpose of this particular comparison was to assess the consistency of the sample collected in 2011 with the samples collected from prior years. Two further comparisons were also made. Estimates for 2008-2009 were compared with estimates for 2010-2011 to determine whether changes in the sampling design implemented in 2010 and 2011 (see Section 3.6 for details) affected the demographic, mental health, and substance use characteristics between the samples. In addition, estimates for 2010 were compared with those for 2011 to examine the impact of random sampling variation occurring under the same sampling design with an increased sample size. Key demographic characteristics included gender, age, race/ethnicity, education, poverty threshold, core-based statistical area (CBSA), and employment status. Mental health characteristics included lifetime and past year depression, depression treatment, mental health treatment, and suicidality measures. Substance use characteristics included past month tobacco and marijuana use and past year alcohol abuse or dependence. Mental health characteristics from the SCID included SMI status and diagnoses of mental health and substance use. Details about these variables can be obtained from the 2011 NSDUH national findings report (CBHSQ, 2012b) and the 2011 NSDUH mental health findings report (CBHSQ, 2012c).

Overall, there appears to be evidence that modifications to the sampling design from 2008 to 2011 had differential effects on the unweighted distributions of several key characteristics in each year's MHSS clinical data. The modifications to the sampling design reduced the oversampling of respondents with high K6 scores and increased the undersampling of respondents with low K6 scores. This had the effect that the sampled distributions were more similar to the population distributions, and the variation in the size of the weights was also reduced. As proven through the weighted descriptive analysis, taking weights into account removed most of these differential effects. However, there were still some significant effects for the characteristics in the weighted data, including the following:

- education when comparing 2008-2010 with 2011;
- past year receiving outpatient mental health treatment when comparing 2008-2010 with 2011, when comparing 2008-2009 with 2010-2011, and when comparing 2010 with 2011;
- past year receiving any mental health treatment when comparing 2008-2010 with 2011;
- Global Assessment of Functioning (GAF) score less than or equal to 50 when comparing 2008-2010 with 2011; and

- any mental illness (AMI) when comparing 2008-2010 with 2011.

When making comparisons of the estimates across different time periods, attention should be paid to whether the compared estimates are correlated with one or more characteristics that have imbalanced distributions across the compared time periods. Differences in estimates across time periods may not reflect a true difference in populations across the compared time periods, but result instead from the fact that the distributions of related characteristics are different (i.e., are imbalanced) in the samples collected from the different time periods. For details, see Liao et al. (2012).

6.3 Evaluating the Overall 2008 to 2012 Sample

Initial descriptive analyses and statistical tests compared key demographic, mental health, and substance use characteristics between the MHSS clinical sample from each individual year and the combined MHSS clinical samples from 2008 through 2012. The purpose of this particular comparison was to assess the consistency of samples collected from different sampling periods.

Key demographic characteristics included age, gender, race/ethnicity, education, poverty threshold, CBSA, and employment status. Mental health characteristics included lifetime and past year depression, depression treatment, mental health service use, suicidality measures, and some mental health characteristics from the SCID, including mental illness status and substance use disorder. Substance use characteristics included past month substance use and past year substance abuse or dependence. Details about these variables can be obtained from the 2011 NSDUH national findings report (CBHSQ, 2012b) and the 2011 NSDUH mental health findings report (CBHSQ, 2012c).

Unweighted descriptive statistics of the demographic, mental health, and substance use measures based on the MHSS clinical sample are shown in Tables 6.1, 6.2, and 6.3, and weighted versions of those descriptive statistics based on both the MHSS clinical sample and the all-adult NSDUH main interview sample are shown in Tables 6.4, 6.5, and 6.6.[42] Included in the descriptive statistics are percentages across the following six sampling periods: (1) 2008, (2) 2009, (3) 2010, (4) 2011, (5) 2012, and (6) the combined 2008-2012. The weighted percentages for the combined 2008-2012 clinical sample used the scaled analysis weights.[43] For comparison purposes, the combined 2008-2012 adult NSDUH main study samples used both the unscaled NSDUH analysis weights and the scaled NSDUH analysis weights with the same scaling factors as were applied to the clinical sample. Statistical tests compared each sampling period from the clinical sample and the adult NSDUH main study sample, as well as the combined 2008-2012 sampling period from the adult NSDUH main study sample with the combined 2008-2012 sampling period from the clinical sample.

[42] To facilitate the data presentation and discussion, all of the tables in this chapter have been grouped at the end of the chapter's text.

[43] Scaling was applied to the sample weights to reduce the variance of estimates and increase statistical efficiency. See Section 5.5 for more details.

6.3.1 Demographic Characteristics

Table 6.1 shows that with unweighted data there was a statistically significant difference between some individual year clinical sample and the combined 2008-2012 clinical sample for some demographic characteristics, including age, education, poverty, population density, and employment status. These differences appeared to be driven by both the initial change in the sampling design after the first two quarters of 2009 and the change in 2010, which allowed a greater proportion of older respondents and persons with lower K6 and WHODAS scores to be sampled for the MHSS clinical study. No significant effect was evident in the tests for gender and Hispanic origin and race.

Table 6.4 indicates that applying weights mitigated most of the differences among the demographic characteristics that appeared in the unweighted data of the clinical sample. That is, corresponding *p* values increased and often became insignificant. The only significant difference remaining was for adults with an education level of less than high school in the 2008 clinical sample. Note that although the 2008-2012 unweighted proportion of this subpopulation was lower than its corresponding 2008 counterpart, the 2008-2012 weighted proportion was higher than its corresponding 2008 counterpart. However, three insignificant differences among the unweighted statistics turned into significant differences after the weights were applied: (1) adults living in a CBSA with 1 million or more people in 2008, (2) adults who were unemployed in the past year in 2008, and (3) adults who were in the "other" category of employment status in 2009. Among these three differences, the weighted proportions of the combined 2008-2012 clinical sample were all relatively higher than their corresponding counterparts in 2008 or 2009.

Table 6.4 also displays weighted percentages derived from the main computer-assisted interviewing (CAI) sample (i.e., the all-adult NSDUH main interview sample). As shown in the last two columns, the 2008-2012 CAI sample distributions based on both the unscaled and scaled weights were generally consistent with their corresponding SCID counterparts without any significant differences, which implies that the revised weighting procedures for the clinical sample (as described in Chapter 5) effectively adjusted the clinical sample distributions of these key demographic characteristics to be consistent with the all-adult NSDUH main study sample distributions. There was only one exception for adults with a poverty level larger than or equal to the 200 percent threshold. The weighted percentage for this group in the CAI sample with scaled weights was slightly smaller than its counterpart in the SCID sample (65.9 vs. 68.6 percent, respectively). A few differences were also found when comparing the 2008-2012 SCID weighted percentages with the individual year CAI weighted percentages. These differences had less of an impact in these analyses because scaled weights were used for analyzing the clinical sample, while unscaled weights were used for analyzing the all-adult NSDUH main study sample.

6.3.2 Mental Health Characteristics

Table 6.2 presents the unweighted percentages of some mental health characteristics in the SCID clinical sample. As shown in the table, significant differences were found in all four mental health categories when comparing an individual year sample with the combined 2008-2012 sample, including major depressive episode (MDE), past year treatment for depression, past year mental health service use, suicidal experiences, and some key variables from the SCID.

After the weights were applied, as shown in Table 6.5, all but one of the significant differences in the unweighted data became insignificant. The only significant difference remaining was for the adults who received inpatient mental health service in the past year in 2011. Also, three insignificant differences in the unweighted data became significant when the weights were applied: (1) adults who received inpatient mental health service in the past year in 2009, (2) adults who received outpatient mental health service in the past year in 2011, and (3) adults who had attempted suicide in the past year in 2010. Although these results may be due to real underlying differences in the prevalence rates being estimated, they also can be due to a combination of the small yearly clinical sample sizes and the small prevalence rates. Tests performed for differences assume the asymptotic normality of the test statistic. This assumption may not be reasonable when there are small sample sizes and prevalence rates. As in Table 6.4, Table 6.5 also displays weighted percentages derived from the CAI sample. No significant difference was found between the 2008-2012 SCID sample and the 2008-2012 CAI sample, which implies that the sample distributions of these two samples were consistent for the compared key characteristics in this table.

6.3.3 Substance Use Characteristics

Table 6.3 indicates that for the unweighted clinical sample, there was a significant difference between some of the individual samples (including clinical samples from 2008, 2009, 2011, and 2012 and the combined 2008-2012 clinical sample) for some substance use characteristics, including past month substance use and past year substance abuse or dependence. However, there were no significant differences between the 2010 clinical estimates and combined 2008-2012 clinical sample. Table 6.6 shows that after applying weights to the data there were no longer any significant differences between any individual year clinical sample estimates and the combined 2008-2012 clinical sample estimates. The weighted distributions of all of these substance use characteristics were consistent among the SCID and CAI samples across different sampling periods.

6.3.4 K6 and WHODAS Total Scores

Unweighted descriptive statistics of the past year K6 total score (i.e., the maximum of the past 30-day K6 total score and the worst month K6 total score) of the SCID samples are shown in Table 6.7, and similar weighted descriptive statistics of the SCID samples and the CAI samples are shown in Table 6.8. The K6 scores were compared between the individual year SCID samples and the CAI samples with the combined 2008-2012 SCID sample by CMH tests of general association (24 *df*) and means (1 *df*). For the unweighted scores, results show that each individual year SCID sample was significantly different from the combined 2008-2012 SCID sample, except for the 2010 SCID sample with respect to the test of means. Weights removed all of the other significant differences for the tests of means. However, all of the general association tests still remained significant for the weighted scores in the SCID sample, which might be caused by sampling design changes on the distribution of past year K6 total scores across years. Both individual year CAI samples and the combined 2008-2012 CAI sample with unscaled or scaled weights showed no differences when compared with the combined 2008-2012 SCID sample with respect to the tests of means. Similar to the individual year SCID samples, all of the general association tests were statistically significant for the weighted scores in the CAI samples when compared with the combined 2008-2012 SCID sample. Weighted descriptive statistics of

the past 30-day K6 total score and the past year K6 total score for all of the sampling periods are given in Table 6.9. Their means were compared between the individual year SCID samples and the combined 2008-2012 SCID sample by the *t* test. All of the means from each individual year SCID sample were found insignificantly different from the means of the combined 2008-2012 SCID sample. In addition, no difference was found for the 2008-2012 CAI sample with unscaled or scaled weights when compared with the 2008-2012 SCID sample.

The weighted frequency distributions of the past year WHODAS total scores for the SCID and CAI samples are given in Table 6.10. The WHODAS scores were also compared between the individual year SCID samples, the individual year CAI samples, and the combined 2008-2012 CAI sample with the combined 2008-2012 SCID sample, respectively, by CMH tests of general association (24 *df*) and means (1 *df*). Similar to the results for the K6 total score, no significant difference was evident among the tests of means with the weighted data, while significant differences were detected among all of the tests of general association except for the 2009 CAI sample.

6.3.5 Summary

Overall, some significant differences were found for some key demographic, mental health, and substance use characteristics when comparing each individual year clinical sample with the combined 2008-2012 clinical sample, which indicates some inconsistencies across unweighted clinical samples from different sampling periods. These differences could be driven by four factors: (1) the initial change in the sampling design after the first two quarters of 2009 and the change in 2010, (2) the small yearly clinical sample, (3) the small yearly prevalence rates of the characteristics, and (4) the real underlying changes of the yearly prevalence rates across different sampling periods in the clinical study.

Applying weights to the data removed most of these differential effects between estimates. However, there were still some significant differences for the characteristics in the weighted data, as follows:

- education level less than high school (in 2008),
- living in a CBSA with 1 million or more people (in 2008),
- unemployed in the past year (in 2008),
- having "other" employment status (in 2009),
- received inpatient mental health service in the past year (in 2011),
- received inpatient mental health service in the past year (in 2009),
- received outpatient mental health service in the past year (in 2011), and
- attempted suicide in the past year (in 2010).

When making comparisons of the estimates across different sampling periods, attention should be paid to whether the compared estimates are correlated with one or more characteristics that have imbalanced distributions across the compared time periods. Differences in estimates across time periods may result from different sample distributions of these characteristics and

thus cannot reflect a true difference in populations across the compared time periods. In addition, no significant test was found between the weighted 2008-2012 clinical sample (with scaled weights) and the weighted 2008-2012 all-adult NSDUH main interview sample (with unscaled weights), which implies that the revised MHSS analysis weights have adjusted the clinical sample well enough to be consistent with the all-adult NSDUH interview study sample.

Table 6.1 Sample Description of SCID Respondents, by Demographic Characteristics: 2008-2012

Characteristic	2008 Unweighted Pct.	2009 Unweighted Pct.	2010 Unweighted Pct.	2011 Unweighted Pct.	2012 Unweighted Pct.	2008-2012 Unweighted Pct.
Total (Sample Size)	1,500	520	516	1,495	1,622	5,653
Age						
18-25	58.7[a]	56.3[a]	24.8[a]	20.9[a]	21.3[a]	34.7
26-34	14.9[a]	14.2[a]	25.8[a]	24.7[a]	22.3	20.6
35-49	17.9[a]	18.8[a]	33.7[a]	31.0[a]	30.1[a]	26.4
50+	8.5[a]	10.6[a]	15.7	23.4[a]	26.3[a]	18.4
Age						
18-30	67.2[a]	63.8[a]	40.3[a]	34.0[a]	33.9[a]	46.1
31+	32.8[a]	36.2[a]	59.7[a]	66.0[a]	66.1[a]	53.9
Gender						
Male	36.5[a]	42.1	40.3	40.5	38.9	39.1
Female	63.5[a]	57.9	59.7	59.5	61.1	60.9
Hispanic Origin and Race						
Not Hispanic or Latino						
White	71.6	72.3	72.1	73.2	72.7	72.5
Black or African American	10.8	10.0	10.3	9.6	10.2	10.2
Other or Multiple Races	7.5	7.1	8.1	7.4	7.6	7.5
Hispanic or Latino	10.1	10.6	9.5	9.8	9.4	9.8
Education						
< High School	14.3[a]	11.7	10.5	9.6[a]	10.3	11.3
High School Graduate	29.3	31.3	26.9	27.3	27.9	28.3
Some College	33.4	31.5	32.0	32.6	30.1[a]	31.9
College Graduate	23.1[a]	25.4	30.6	30.5[a]	31.7[a]	28.4
Poverty[1]						
< 100% Threshold	18.1	17.7	11.9[a]	15.2	16.8	16.3
100-199% Threshold	23.4	21.7	21.6	22.1	20.6	21.9
≥ 200% Threshold	58.4[a]	60.6	66.5[a]	62.8	62.6	61.7
Population Density						
CBSA ≥ 1M	39.1	40.2	43.4	38.3	40.4	39.7
250K ≤ CBSA < 1M	25.5	26.0	23.3	26.0	22.7[a]	24.7
CBSA < 250K	26.3	27.3	24.0	26.7	28.7	27.0
Non-CBSA, Urban	3.6	2.1	2.7	2.9	1.7[a]	2.7
Non-CBSA, Rural	5.4	4.4	6.6	6.2	6.5	5.9
Employment Status						
Full Time	49.5	49.2	55.0	49.9	51.9	50.8
Part Time	23.4[a]	21.2	15.1[a]	16.7[a]	16.5[a]	18.7
Unemployed	6.5	8.7	9.3	6.7	6.6	7.0
Other[2]	20.7[a]	21.0	20.5	26.8[a]	25.0	23.5

CAI = computer-assisted interviewing; CBSA = core-based statistical area; K = thousand; K6 = Kessler-6, a 6-item psychological distress scale; M = million; NSDUH = National Survey on Drug Use and Health; pct. = percent; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. The respondent characteristics in this table are based on data collected from the NSDUH CAI interview.

[a] Difference between this estimate and the corresponding estimate from 2008-2012 is statistically significant at the .05 level.

[1] U.S. census poverty threshold. Persons aged 18 to 22 in a college dormitory were excluded from the analysis.

[2] The other employment category includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 6.2 Sample Description of SCID Respondents, by Mental Health Characteristics: 2008-2012

Characteristic	2008 Unweighted Pct.	2009 Unweighted Pct.	2010 Unweighted Pct.	2011 Unweighted Pct.	2012 Unweighted Pct.	2008-2012 Unweighted Pct.
Total (Sample Size)	1,500	520	516	1,495	1,622	5,653
MDE[1]						
Lifetime/Not Past Year MDE	11.5	11.1	12.7	10.3	10.5	11.0
Past Year MDE						
Without Impairment	7.3[a]	6.6	6.3	5.3	4.8[a]	5.9
With Impairment	14.2	12.0	15.6	12.1[a]	15.2	13.9
No Lifetime/Past Year MDE	66.9[a]	70.3	65.4	72.3[a]	69.5	69.3
Past Year Treatment for Depression[2]						
Any Treatment	63.6[a]	56.3[a]	74.1	73.0	72.1	68.8
Saw or Talked to Health Professional	60.0	51.0[a]	67.9	69.1	67.2	64.3
Prescription Medication	41.3[a]	44.8	64.3[a]	56.8	56.3	52.1
Past Year Mental Health Service Use[3]						
Any Treatment	25.6	25.3	32.2[a]	24.6	27.2	26.4
Outpatient	15.3	16.4	18.1	13.4	15.1	15.1
Inpatient	1.0	0.6	1.7	0.7[a]	1.7[a]	1.2
Prescription Medication	20.7	20.6	27.1[a]	20.5	23.9	22.1
Suicidal Experiences[4]						
Had Thoughts of Suicide	13.7[a]	10.0	10.5	9.7	9.7[a]	10.9
Made Plans for Suicide	4.5[a]	3.1	2.9	2.9	2.9	3.3
Attempted Suicide	2.1[a]	0.8	1.0	1.0	1.3	1.4
SCID Variables[5]						
Any Mental Illness[6]	43.8[a]	44.4[a]	37.4	33.9[a]	33.9[a]	37.8
Serious Mental Illness	11.5	10.2	12.4	11.0	11.4	11.3
Moderate Mental Illness	11.3	10.2	10.9	10.8	11.0	10.9
Mild Mental Illness	21.1[a]	24.0[a]	14.1	12.1[a]	11.5[a]	15.6
No Mental Illness	56.2[a]	55.6[a]	62.6	66.1[a]	66.1[a]	62.2
Substance Use Disorder	16.6[a]	16.2	13.2	10.9[a]	10.1[a]	12.9

CAI = computer-assisted interviewing; GAF = Global Assessment of Functioning; K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; NSDUH = National Survey on Drug Use and Health; pct. = percent; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. Most of the respondent characteristics in this table are based on data collected from the NSDUH CAI interview, except for the mental illness and substance use disorder status. These characteristics were determined based on the SCID data. The SCID was the diagnostic tool used for the clinical interview follow-up study to the main NSDUH study.

[a] Difference between this estimate and the corresponding estimate from 2008-2012 is statistically significant at the .05 level.

[1] Respondents with unknown past year MDE data were excluded.
[2] Among those with MDE. "Any Treatment" includes categories either "Saw or Talked to Health Professional" or "Prescription Medication."
[3] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.
[4] Respondents with unknown suicide information were excluded.
[5] Variables from this section were based on the SCID clinical interview, while other characteristic variables in Tables 6.1 to 6.10 were based on the NSDUH main interview.
[6] Three categories of the level of mental illness severity based on functional impairment are defined based on the SCID disorder diagnosis and GAF scales: mild mental illness, moderate mental illness, and serious mental illness. Any mental illness includes persons in any of the three categories. For more details, see Table 1.1 in Chapter 1.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 6.3 Sample Description of SCID Respondents, by Substance Use Measures: 2008-2012

Characteristic	2008 Unweighted Pct.	2009 Unweighted Pct.	2010 Unweighted Pct.	2011 Unweighted Pct.	2012 Unweighted Pct.	2008-2012 Unweighted Pct.
Total (Sample Size)	1,500	520	516	1,495	1,622	5,653
Past Month Substance Use						
Illicit Drugs[1]	15.9[a]	18.8[a]	15.1	13.0	10.8[a]	13.9
Marijuana	13.3[a]	15.0[a]	11.4	10.5	8.7[a]	11.2
Illicit Drugs Other Than Marijuana[1]	6.7[a]	7.9[a]	6.6	4.2[a]	3.7[a]	5.3
Tobacco Products[2]	40.3[a]	41.0	38.6	35.5	34.3[a]	37.2
Alcohol	61.9	65.0	62.8	60.7	61.7	61.9
Past Year Substance Abuse or Dependence						
Substance Use Disorder[3]	22.3[a]	20.8[a]	13.6	13.4[a]	13.8[a]	16.6
Alcohol Use Disorder	18.5[a]	17.7[a]	12.0	10.5[a]	11.2[a]	13.6
Illicit Drug Use Disorder	7.9[a]	7.7	4.5	4.7[a]	4.6[a]	5.8
No Substance Use Disorder	77.7[a]	79.2[a]	86.4	86.6[a]	86.2[a]	83.4

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; NSDUH = National Survey on Drug Use and Health; pct. = percent; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. The respondent characteristics in this table are based on data collected from the NSDUH CAI interview.

[a] Difference between this estimate and the corresponding estimate from 2008-2012 is statistically significant at the .05 level.

[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other Than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically, including data from original methamphetamine questions but not including new methamphetamine items added in 2005 and 2006.

[2] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.

[3] Substance Use Disorder is defined as meeting criteria for illicit drug or alcohol dependence or abuse. Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 6.4 Weighted Sample Description of SCID and CAI Respondents, by Demographic Characteristics: 2008-2012

	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
Characteristic	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct. (unscaled)	Wt. Pct. (scaled)
Total (Numbers in Thousands)	224,923	224,923	227,207	227,207	229,273	229,273	232,625	232,625	235,124	235,124	231,890	229,830	231,890
Age													
18-25	14.6	14.6	14.8	14.8	14.9	14.9	14.7	14.7	14.7	14.7	14.7	14.7	14.7
26-34	18.6	15.8	19.1	15.9	14.5	15.9	16.6	15.7	16.6	15.7	16.6	15.8	15.7
35-49	25.8	28.5[a]	24.7	27.8	28.5	27.1	25.6	26.5	25.1	26.0	25.8	27.2	26.7
50+	41.0	41.0	41.5	41.5	42.1	42.1	43.1	43.1	43.6	43.6	42.8	42.3	42.8
Age													
18-30	25.6	23.9	26.3	23.8	24.3	24.0	23.6	23.7	24.0	23.7	24.2	23.8	23.8
31+	74.4	76.1	73.7	76.2	75.7	76.0	76.4	76.3	76.0	76.3	75.8	76.2	76.2
Gender													
Male	48.3	48.3	48.3	48.3	48.4	48.4	48.1	48.1	48.1	48.1	48.2	48.2	48.2
Female	51.7	51.7	51.7	51.7	51.6	51.6	51.9	51.9	51.9	51.9	51.8	51.8	51.8
Hispanic Origin and Race													
Not Hispanic or Latino													
White	68.8	68.8	68.4	68.4	68.0	68.0	66.7	66.7	66.3	66.3	67.1	67.6	67.1
Black or African American	11.3	11.3	11.6	11.6	11.6	11.6	11.5	11.5	11.6	11.6	11.5	11.5	11.5
Other or Multiple Races	6.4	6.4	6.3	6.3	6.4	6.4	7.2	7.2	7.4	7.4	7.0	6.7	7.0
Hispanic or Latino	13.5	13.5	13.7	13.7	13.9	13.9	14.6	14.6	14.8	14.8	14.4	14.1	14.4
Education													
< High School	7.6[a]	15.6[a]	14.2	15.4	13.8	15.3[a]	12.8	14.2	14.8	14.6	13.1	15.0	14.7
High School Graduate	32.3	31.0	28.1	30.7	31.3	30.2	28.3	30.0	28.8	29.7	29.4	30.3	30.0
Some College	30.3	25.5	23.4	25.4	25.9	25.8	29.7	26.5	25.2	26.6	27.4	26.0	26.3
College Graduate	29.8	27.9	34.3	28.6	29.1	28.7	29.3	29.3	31.2	29.2	30.2	28.7	29.0
Poverty[1]													
< 100% Threshold	11.9	11.3	9.3	12.2	9.5	13.0	13.2	14.1	13.3	15.9[a]	12.4	13.3	14.2
100-199% Threshold	14.3	18.6	19.6	19.4	16.9	20.8	19.4	20.6	20.7	19.4	19.0	19.8	19.9
≥ 200% Threshold	73.7	70.1	71.1	68.3	73.6	66.2	67.3	65.3[a]	66.0	64.8[a]	68.6	66.9	65.9[a]
Population Density													
CBSA ≥ 1M	42.7[a]	51.9	50.1	51.7	56.0	52.1	50.0	51.7	52.1	51.9	50.7	51.8	51.8
250K ≤ CBSA < 1M	24.7	21.4	23.7	21.5	17.9	21.7	24.7	22.6	20.1	21.9	22.1	21.8	22.0
CBSA < 250K	23.1	20.4	17.5	20.8	19.5	20.0	19.7	19.5	21.8	20.0	20.7	20.1	19.9
Non-CBSA, Urban	2.8	2.0	0.9	1.8	1.3	2.1[a]	1.6	1.8	1.2	1.6	1.5	1.8	1.8
Non-CBSA, Rural	6.8	4.3	7.8	4.2	5.2	4.2	4.0	4.6	4.8	4.7	4.9	4.4	4.5
Employment Status													
Full Time	57.8	54.4[a]	60.9	50.5	53.0	49.8	47.3	49.7	50.8	50.0	51.1	50.9	50.4
Part Time	12.6	13.5	14.6	14.0	11.9	14.5	13.7	13.9	13.2	13.9	13.2	14.0	14.0
Unemployed	2.6[a]	4.0[a]	5.2	6.6	8.7	6.5	7.7	5.8	6.0	5.8	6.5	5.7	5.7
Other[2]	27.0	28.2	19.3[a]	29.0	26.4	29.3	31.4	30.5	30.0	30.3	29.2	29.4	29.9

CAI = computer-assisted interviewing; CBSA = core-based statistical area; K = thousand; K6 = Kessler-6, a 6-item psychological distress scale; M = million; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule; wt. pct. = weighted percent.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. The respondent characteristics in this table are based on data collected from the NSDUH CAI interview.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

[a] Difference between this estimate and the corresponding estimate from the 2008-2012 SCID data is statistically significant at the .05 level.

[1] U.S. census poverty threshold. Persons aged 18 to 22 in a college dormitory were excluded from the analysis.

[2] The other employment category includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 6.5 Weighted Sample Description of SCID and CAI Respondents, by Mental Health Characteristics: 2008-2012

	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
Characteristic	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct. (unscaled)	Wt. Pct. (scaled)
Total (Numbers in Thousands)	224,923	224,923	227,207	227,207	229,273	229,273	232,625	2 32,625	235,124	235,124	231,890	229,830	231,890
MDE[1]													
Lifetime/Not Past Year MDE	6.4	6.4	6.5	6.4	5.8	5.9	6.1	6.1	6.3	6.3	6.2	6.2	6.2
Past Year MDE													
Without Impairment	2.4	2.4	3.0	2.6	2.2	2.6	2.8	2.4	1.8	2.4	2.3	2.5	2.4
With Impairment	4.4	4.4	3.7	4.0	4.7	4.2	3.8	4.2	5.0	4.5	4.4	4.3	4.3
No Lifetime/Past Year MDE	86.8	86.8	86.8	87.0	87.4	87.3	87.3	87.3	86.8	86.8	87.1	87.0	87.1
Past Year Treatment for Depression[2]													
Any Treatment	70.5	69.0	76.4	64.3	67.6	68.2	66.8	68.1	70.4	68.0	69.0	67.5	68.0
Saw or Talked to Health Professional	61.4	63.5	75.0	58.9	65.0	62.7	62.2	62.2	66.0	62.7	64.4	62.0	62.5
Prescription Medication	50.5	52.9	65.4	50.0	60.9	52.1	48.3	51.7	53.7	51.6	52.9	51.7	51.8
Past Year Mental Health Service Use[3]													
Any Treatment	15.9	13.4	12.3	13.4	14.7	13.8	12.0	13.6	13.9	14.5	13.5	13.8	13.9
Outpatient	8.1	6.8	7.8	6.4	6.7	6.6	4.9[a]	6.7	6.7	6.6	6.3	6.6	6.6
Inpatient	0.4	0.9	0.2[a]	0.8	1.4	0.8	0.2[a]	0.8	1.0	0.8	0.7	0.8	0.8
Prescription Medication	13.3	11.4	10.9	11.3	11.3	11.7	10.3	11.5	11.6	12.4	11.3	11.7	11.8
Suicidal Experiences[4]													
Had Thoughts of Suicide	3.7	3.7	3.7	3.7	3.8	3.8	3.7	3.7	3.8	3.9	3.8	3.8	3.8
Made Plans for Suicide	1.0	1.0	1.3	1.0	1.3	1.1	0.8	1.0	1.2	1.1	1.0	1.1	1.1
Attempted Suicide	0.2	0.5	0.2	0.5	0.1[a]	0.5	0.3	0.5	0.6	0.6[a]	0.4	0.5	0.5
SCID Variables[5]													
Any Mental Illness[6]	19.8	--	20.3	--	17.7	--	17.1	--	18.1	--	18.0	--	--
Serious Mental Illness	4.7	--	5.6	--	3.2	--	3.7	--	4.1	--	3.9	--	--
Moderate Mental Illness	4.2	--	5.0	--	4.0	--	5.9	--	4.9	--	5.1	--	--
Mild Mental Illness	10.9	--	9.6	--	10.5	--	7.4	--	9.2	--	9.0	--	--
No Mental Illness	80.2	--	79.7	--	82.3	--	82.9	--	81.9	--	82.0	--	--
Substance Use Disorder	7.3	--	6.3	--	9.2	--	7.0	--	8.1	--	7.7	--	--

CAI = computer-assisted interviewing; GAF = Global Assessment of Functioning; K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule; wt. pct. = weighted percent.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. Most of the respondent characteristics in this table are based on data collected from the NSDUH CAI interview, except for the mental illness and substance use disorder status. These characteristics were determined based on the SCID data. The SCID was the diagnostic tool used for the clinical interview follow-up study to the main NSDUH survey.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

[a] Difference between this estimate and the corresponding estimate from the 2008-2012 SCID data is statistically significant at the .05 level.

[1] Respondents with unknown past year MDE data were excluded.

[2] Among those with MDE. "Any Treatment" includes categories either "Saw or Talked to Health Professional" or "Prescription Medication."

[3] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.

[4] Respondents with unknown suicide information were excluded.

[5] Variables from this section were based on the SCID clinical interview, while other characteristic variables in Tables 6.1 to 6.10 were based on the NSDUH main interview.

[6] Three categories of the level of mental illness severity based on functional impairment are defined based on the SCID disorder diagnosis and GAF scales: mild mental illness, moderate mental illness, and serious mental illness. Any mental illness includes persons in any of the three categories. For more details, see Table 1.1 in Chapter 1.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 6.6 Weighted Sample Description of SCID and CAI Respondents, by Substance Use Measures: 2008-2012

	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
Characteristic	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct.**	**Wt. Pct. (unscaled)**	**Wt. Pct. (scaled)**
Total (Numbers in Thousands)	224,923	224,923	227,207	227,207	229,273	229,273	232,625	232,625	235,124	235,124	231,890	229,830	231,890
Past Month Substance Use													
Illicit Drugs[1]	7.4	7.9	8.4	8.6	10.0	8.8	9.7	8.6	8.2	9.1	8.9	8.6	8.7
Marijuana	6.3	6.1	7.1	6.6	8.5	6.8	8.0	6.9	6.1	7.3	7.2	6.7	6.9
Illicit Drugs Other Than Marijuana[1]	2.4	3.3	5.3	3.6	3.2	3.5	2.4	3.0	2.9	3.4	2.8	3.3	3.3
Tobacco Products[2]	28.8	30.3	23.9	29.5	28.1	29.2	30.8	28.3	28.7	28.6	29.2	29.2	28.8
Alcohol	57.7	55.7	61.1	55.9	56.6	55.9	58.5	55.9	57.3	56.3	57.8	55.9	56.0
Past Year Substance Abuse or Dependence													
Alcohol or Drug Substance Use Disorder[3]	9.0	9.1	9.0	9.2	7.9	8.9	8.9	8.1	8.8	8.8	8.7	8.8	8.6
Alcohol Use Disorder	7.9	7.7	6.9	7.8	7.4	7.4	7.2	6.8	7.6	7.2	7.4	7.3	7.1
Illicit Drug Use Disorder	2.4	2.6	3.6	2.7	1.6	2.6	2.8	2.3	2.9	2.7	2.6	2.6	2.5
No Substance Use Disorder	91.0	90.9	91.0	90.8	92.1	91.1	91.1	91.9	91.2	91.2	91.3	91.2	91.4

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule; wt. pct. = weighted percent.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. The respondent characteristics in this table are based on data collected from the NSDUH CAI interview.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

[a] Difference between this estimate and the corresponding estimate from the 2008-2012 SCID data is statistically significant at the .05 level.

[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other Than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically, including data from original methamphetamine questions but not including new methamphetamine items added in 2005 and 2006.

[2] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.

[3] Substance Use Disorder is defined as meeting criteria for illicit drug or alcohol dependence or abuse. Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 6.7 Sample Description of SCID Respondents, by Past Year K6 Score Frequency Distribution: 2008-2012

	2008	2009	2010	2011	2012	2008-2012
Past Year K6 Score	**Unweighted Pct.**	**Unweighted Pct.**	**Unweighted Pct.**	**Unweighted Pct.**	**Unweighted Pct.**	**Unweighted Pct.**
0	2.4	4.0	10.7	11.4	11.1	8.2
1	1.1	1.9	6.2	7.6	6.4	4.9
2	1.7	1.9	6.8	7.0	7.3	5.2
3	1.3	3.5	4.5	6.6	5.5	4.4
4	3.6	4.8	6.8	5.4	5.6	5.1
5	3.1	2.9	3.5	5.6	4.7	4.3
6	3.8	5.0	4.3	3.9	3.8	4.0
7	3.4	5.2	4.5	3.9	3.8	3.9
8	6.7	8.1	3.1	3.3	4.1	4.8
9	7.4	4.4	2.9	3.9	3.0	4.5
10	7.7	5.6	1.2	2.5	4.0	4.5
11	6.1	6.9	3.9	4.6	3.0	4.7
12	11.2	9.8	3.3	5.5	5.0	7.1
13	7.4	7.5	6.4	3.3	3.8	5.2
14	5.3	3.8	3.9	3.5	3.9	4.1
15	5.6	6.3	3.7	3.1	3.7	4.3
16	4.1	3.7	4.1	2.7	3.6	3.6
17	2.9	1.7	3.7	3.1	3.6	3.1
18	5.0	4.0	5.2	3.3	4.7	4.4
19	1.5	2.1	2.1	2.4	2.1	2.0
20	1.8	1.5	2.1	1.4	2.1	1.8
21	1.3	1.5	1.2	1.0	0.6	1.0
22	0.9	1.2	0.8	1.2	1.4	1.1
23	0.9	0.2	1.6	0.8	0.5	0.8
24	3.8	2.3	3.9	2.7	2.8	3.1
Total (Sample Size)	1,500	520	516	1,495	1,622	5,653

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; pct. = percent; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. Past year K6 total scores in this table are based on data collected from the NSDUH CAI interview.

NOTE: K6SCMAX was used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2012.

Table 6.8 Weighted Sample Description of SCID and CAI Respondents, by Past Year K6 Score Frequency Distribution: 2008-2012

Past Year K6 Score	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct. (unscaled)	Wt. Pct. (scaled)
0	15.2	22.0	20.7	22.9	25.1	23.3	22.7	23.9	20.2	23.5	21.2	23.1	23.4
1	9.8	10.8	3.8	10.3	11.2	10.5	12.3	10.9	9.9	10.4	10.7	10.6	10.6
2	9.9	10.7	7.9	10.9	9.7	10.4	10.4	11.1	12.7	10.7	11.0	10.8	10.8
3	9.4	9.5	15.5	8.7	9.4	9.1	11.2	8.7	8.2	8.9	9.8	9.0	8.9
4	8.8	7.9	5.9	7.8	6.8	7.9	6.2	7.6	9.8	7.6	7.8	7.8	7.7
5	7.3	6.1	6.6	6.1	5.1	5.8	7.2	5.7	5.4	6.2	6.2	6.0	5.9
6	6.7	5.3	4.5	5.5	5.1	5.2	3.9	5.4	3.6	5.4	4.3	5.4	5.4
7	6.9	4.1	11.4	4.0	6.3	3.9	3.9	3.7	5.2	3.7	5.3	3.9	3.8
8	4.2	3.3	3.1	3.5	2.0	3.4	2.4	3.2	4.5	3.2	3.3	3.3	3.2
9	4.4	2.7	0.7	2.6	3.2	2.9	2.5	2.7	2.9	2.6	2.9	2.7	2.7
10	2.2	2.6	3.2	2.4	0.9	2.4	1.8	2.4	2.2	2.4	1.9	2.4	2.4
11	1.6	2.0	1.5	2.3	2.5	2.1	2.4	2.1	1.8	2.1	2.1	2.1	2.1
12	3.8	2.7	5.9	2.8	1.2	2.8	2.7	2.5	2.8	2.6	2.8	2.7	2.6
13	2.2	1.6	1.7	1.6	2.0	1.5	1.3	1.5	1.8	1.6	1.7	1.6	1.6
14	0.8	1.4	0.8	1.4	1.2	1.3	1.5	1.3	1.5	1.3	1.3	1.3	1.3
15	1.4	1.1	1.7	1.2	0.9	1.2	1.4	1.0	1.0	1.1	1.2	1.1	1.1
16	1.0	1.1	0.9	1.1	1.8	1.1	1.2	1.1	1.4	1.0	1.3	1.1	1.1
17	0.8	1.0	0.4	0.9	1.1	0.8	1.1	0.9	0.8	0.9	0.9	0.9	0.9
18	1.3	1.4	1.3	1.2	1.2	1.5	1.2	1.2	1.5	1.5	1.4	1.4	1.4
19	0.4	0.5	0.6	0.6	0.4	0.6	0.7	0.6	0.6	0.7	0.6	0.6	0.6
20	0.6	0.5	0.7	0.4	1.2	0.5	0.4	0.5	0.6	0.6	0.6	0.5	0.5
21	0.5	0.4	0.4	0.4	0.2	0.3	0.3	0.3	0.1	0.4	0.2	0.4	0.3
22	0.1	0.3	0.2	0.3	0.2	0.3	0.3	0.3	0.5	0.3	0.3	0.3	0.3
23	0.1	0.2	0.0	0.2	0.3	0.2	0.2	0.2	0.3	0.3	0.2	0.2	0.3
24	0.8	0.8	0.7	0.9	1.0	1.0	0.8	1.1	0.8	1.0	0.8	1.0	1.0
Total (Numbers in Thousands)	224,923	224,923	227,207	227,207	229,273	229,273	232,625	232,625	235,124	235,124	231,890	229,830	231,890

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule; wt. pct. = weighted percent.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. Past year K6 total scores in this table are based on data collected from the NSDUH CAI interview.

NOTE: K6SCMAX was used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 6.9 Weighted Sample Description of SCID and CAI Respondents, by K6 Scores: 2008-2012

	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
K6 Score	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE)**	**Mean (SE) (unscaled)**	**Mean (SE) (scaled)**
Past Month Total Score	4.11 (0.22)	3.77 (0.03)	4.27 (0.47)	3.79 (0.03)	3.53 (0.25)	3.79 (0.04)	3.66 (0.14)	3.70 (0.03)	3.92 (0.13)	3.83 (0.04)	3.81 (0.08)	3.77 (0.02)	3.77 (0.02)
Past Year Total Score	5.42 (0.30)	4.90 (0.04)	5.37 (0.50)	4.89 (0.04)	4.79 (0.29)	4.88 (0.04)	4.74 (0.17)	4.78 (0.04)	5.07 (0.14)	4.91 (0.04)	4.97 (0.10)	4.87 (0.02)	4.86 (0.02)

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; SE = standard error; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. Past month and Past year K6 total scores in this table are based on data collected from the NSDUH CAI interview.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

[a] Difference between this estimate and the corresponding estimate from the 2008-2012 SCID data is statistically significant at the .05 level.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table 6.10 Weighted WHODAS Past Year Total Score Frequency Distribution of SCID and CAI Respondents, by Sampling Period: 2008-2012

WHODAS Total Score	2008		2009		2010		2011		2012		2008-2012		
	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	SCID	CAI	
	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct.	Wt. Pct. (unscaled)	Wt. Pct. (scaled)
0	47.4	46.5	38.8	45.6	49.7	46.0	46.0	47.1	43.4	45.8	45.5	46.2	46.4
1	5.8	7.0	2.3	7.2	7.6	7.1	7.3	6.9	6.3	7.3	6.6	7.1	7.1
2	3.5	6.9	11.6	7.1	7.8	7.0	9.0	7.2	9.3	6.6	8.4	7.0	6.9
3	11.3	5.7	10.6	5.9	5.4	5.8	5.2	5.7	5.7	5.6	6.3	5.7	5.7
4	7.9	4.7	4.8	5.1	4.0	4.8	4.0	4.7	3.9	4.9	4.4	4.9	4.8
5	3.0	4.3	8.8	4.1	4.3	4.4	4.0	4.2	5.6	4.1	4.7	4.2	4.2
6	4.1	4.0	2.1	3.8	2.8	3.6	3.4	3.5	3.6	3.7	3.4	3.7	3.7
7	2.9	3.0	6.2	3.2	1.6	3.0	3.4	2.8	3.3	3.3	3.2	3.1	3.0
8	1.7	3.3	2.8	3.3	2.8	3.4	3.9	3.5	3.4	3.3	3.3	3.4	3.4
9	1.1	2.4	0.9	2.2	2.1	2.3	2.0	2.2	1.4	2.1	1.6	2.2	2.2
10	1.2	2.1	2.1	1.9	1.0	2.1	2.1	1.8	2.7	2.2	2.0	2.0	2.0
11	0.8	1.6	1.3	1.7	1.6	1.7	1.4	1.5	2.0	1.6	1.6	1.6	1.6
12	1.9	1.3	0.9	1.6	1.3	1.3	1.7	1.4	1.5	1.5	1.6	1.4	1.4
13	1.9	1.2	1.1	1.2	1.4	1.3	1.3	1.2	1.3	1.2	1.4	1.2	1.2
14	1.8	0.9	1.1	1.2	2.1	1.1	0.8	1.1	1.0	1.1	1.2	1.1	1.1
15	0.8	0.9	0.8	0.9	0.8	0.9	1.0	1.0	1.2	1.0	1.0	1.0	1.0
16	0.7	1.2	0.9	1.0	0.4	0.9	1.2	1.0	1.0	1.2	0.9	1.1	1.1
17	0.2	0.6	0.3	0.5	0.1	0.7	0.4	0.7	0.6	0.7	0.4	0.6	0.7
18	0.4	0.4	0.3	0.6	0.4	0.5	0.6	0.6	1.1	0.7	0.7	0.6	0.6
19	0.3	0.4	0.3	0.4	0.6	0.4	0.3	0.4	0.6	0.4	0.4	0.4	0.4
20	0.4	0.3	0.1	0.3	0.6	0.4	0.2	0.3	0.3	0.3	0.3	0.3	0.3
21	0.2	0.3	1.5	0.3	0.3	0.3	0.4	0.3	0.3	0.3	0.4	0.3	0.3
22	0.5	0.3	0.2	0.3	0.3	0.3	0.2	0.2	0.2	0.3	0.2	0.3	0.3
23	0.2	0.2	0.2	0.2	0.8	0.2	0.1	0.3	0.3	0.2	0.3	0.2	0.2
24	0.2	0.3	0.0	0.3	0.4	0.4	0.2	0.4	0.2	0.4	0.2	0.4	0.4
Total Population (Numbers in Thousands)	224,923	224,923	227,207	227,207	229,273	229,273	232,625	232,625	235,124	235,124	231,890	229,830	231,890

CAI = computer-assisted interviewing; K6 = Kessler-6, a 6-item psychological distress scale; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SCID = Structured Clinical Interview for DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders*, 4th edition); SDS = Sheehan Disability Scale; WHODAS = 8-item World Health Organization Disability Assessment Schedule; wt. pct. = weighted percent.

NOTE: This table does not include any clinical interview (interviews were conducted using the SCID) respondent who had missing values for all K6 and WHODAS/SDS item scores in the NSDUH questionnaire. These respondents were not included in any of the model development analyses. WHODAS past year total scores in this table are based on data collected from the NSDUH CAI interview.

NOTE: For each individual year, the MHSS or the NSDUH annual analysis weight was used to obtain the weighted estimates. For 2008-2012, the rescaled MHSS, the scaled NSDUH, or the unscaled NSDUH analysis weight for the combined 5-year data was used to obtain the weighted estimates. For 2008, MHSAMPWT was used to obtain the NSDUH estimates for the single year data as well as for the combined 5-year data.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

7. Estimation of Mental Illness for the 2008 to 2012 NSDUHs

7.1 Background

The primary objective of the Mental Health Surveillance Study (MHSS) is to produce annual model-based estimates of serious mental illness (SMI) prevalence among the civilian, noninstitutionalized, adult U.S. population using data from the National Survey on Drug Use and Health (NSDUH). These estimates must be of sufficient accuracy, and the examination of trends over time should be possible. Secondary objectives include the estimation of other categories of mental illness, such as any mental illness (AMI), moderate mental illness (MMI), and low (mild) mental illness (LMI).

Although the ideal way to estimate SMI in NSDUH would be to administer a clinical diagnostic interview annually to all 45,000 adult respondents, this approach is not feasible because of constraints on the NSDUH interview time and the need for trained mental health clinicians to conduct the interviews. Therefore, the approach adopted by the Substance Abuse and Mental Health Services Administration (SAMHSA) was to administer a psychological diagnostic interview to a subsample of adult respondents who had completed the NSDUH interview as part of a clinical study from which model-derived estimates of SMI and AMI from the main NSDUH adult samples for each year from 2008 to 2011 were produced.

In 2008, the first year of the MHSS, approximately 1,500 NSDUH respondents participated in the clinical follow-up. The sample collected in 2008 was used to develop prediction models. Specifically, a weighted logistic regression model was fit on the clinical diagnostic data collected in 2008. The dependent variable for the model was whether or not the respondent had a diagnosis of SMI based on the clinical diagnostic interview. The predictor variables were mental health-related items collected in the main NSDUH interview. The model was used to produce a predicted probability of having SMI for each of the respondents to the clinical interview. A cut point was established among the predicted probabilities such that if adults with probabilities at or above the cut point were predicted to have SMI and the rest were not, the weighted number of false positives (adults not diagnosed to have SMI but predicted to have SMI) would come as close as possible to equaling the weighted number of false negatives (adults diagnosed to have SMI but not predicted to have SMI). Because the predictor variables in the model were variables collected on the NSDUH main interview, a probability of having SMI could be predicted for every NSDUH adult respondent using the estimated model parameters. Applying the cut point (determined from the clinical sample) on the predicted probabilities estimated in the NSDUH adult sample, each NSDUH adult respondent was classified as having or not having SMI. This dichotomous variable was then used to compute prevalence estimates of SMI for adults.

For comparability in estimating trends, the 2009, 2010, and 2011 SMI estimates were computed using the method described above. In addition, the estimated logistic model for SMI based on 2008 clinical data was used to determine a cut point for AMI (equalizing the weighted

AMI false positive and false negative as closely as possible in the 2008 clinical sample), which in turn was used to predict the AMI status for NSDUH main interview adult respondents from 2008 to 2011.

Although SAMHSA used the estimation methods based on the 2008 clinical follow-up data for producing annual estimates of SMI, the clinical data collection continued from 2009 to 2012. That is, MHSS clinical interview samples of roughly 500 cases in 2009 and 2010 were collected, followed by 1,500 cases in both 2011 and 2012. After 2012, the annual clinical interviews were discontinued. The data collected from 2008 to 2012 were used to determine whether the 2008 estimation methods should be revised; revisions to the methods would only be implemented if it could be shown that more accurate estimates of SMI could be produced.

Based on the analyses of the nearly 5,500 clinical interviews from 2008 through 2012, it was determined that changes to the 2008 estimation methods would produce considerably more accurate estimates of mental illness. As a result, the 2012 estimates were based on the revised estimation procedures. Also, previously published estimate of mental illness for 2008 through 2011 were revised. Although the clinical sample was discontinued after 2012, the 2012 estimation methods will be used to produce model-based estimates of mental illness annually.

This chapter provides a general overview of the estimation methodology used to produce the prevalence estimates of mental illness among adults aged 18 or older and describes the revisions to the estimation methodology that were implemented in 2012. Specifically, revisions to the estimation methodology included the following:

- the revision of the weights used in fitting the model in the clinical sample (the revisions to the model weights were described in detail in Chapter 5 of this report and are therefore not discussed in detail in this chapter);
- the addition of covariates into the logistic model for SMI to improve the accuracy of SMI predictions and to remove biases in prevalence estimates at the subpopulation level, particularly within age groups; and
- the use of the entire clinical sample from 2008 to 2012 in fitting the model for SMI.

Section 7.2 describes the 2008 estimation method. Section 7.3 provides an overview of the investigation into possible improvements of the 2008 model using the revised weights. A more detailed description of the methodological studies that led to the 2012 estimation methodology can be found elsewhere (Center for Behavioral Health Statistics and Quality [CBHSQ], in press a). Section 7.4 compares SMI and AMI prevalence estimates computed using the 2008 and 2012 estimation methods with their respective weighting schemes. Section 7.5 discusses some caveats of NSDUH model-based mental illness prevalence estimates.

7.2 2008 Estimation Method

As described previously, in 2008, a randomly selected subsample of approximately 1,500 adults who had completed the NSDUH main interview in English was recruited for a follow-up clinical interview consisting of a diagnostic assessment for mental disorders. In addition, as described in greater detail in the first three chapters, a split-sample design was incorporated into

the full 2008 NSDUH data collection in order to determine the optimal scale for measuring functional impairment in NSDUH. Roughly half of the adult respondents were assigned to receive an abbreviated eight-item version of the World Health Organization Disability Assessment Schedule (WHODAS) (Novak et al., 2010), and the other half were assigned to receive the Sheehan Disability Scale (SDS) (Leon et al., 1997). The half sample that received the WHODAS is referred to as the 2008A sample in this report, and the half sample that received the SDS is referred to as the 2008B sample.

Weighted logistic models that predicted SMI were developed for each half sample using the data from the subsample of clinical interview respondents. The short scales (the Kessler-6 [K6] in combination with the WHODAS for the 2008A subsample or the K6 in combination with the SDS for the 2008B subsample) were used as predictors in the weighted logistic models of mental illness assessed via the clinical interviews. The model parameter estimates then were used to predict SMI in the 2008 NSDUH full adult sample based on their responses to the short scale questions.

Based on an analysis of the 2008 MHSS data, it was determined that the WHODAS was the better predictor of SMI and that this scale would be used in combination with the K6 scale to predict SMI. It also was decided that the WHODAS would continue to be administered as the sole impairment scale in the 2009 and subsequent NSDUHs (Aldworth et al., 2009). Therefore, the remainder of this section focuses on the model that was fit on the half sample that was assigned to the WHODAS. This model is called the "2008 (WHODAS) model." A similar model was constructed in the 2008 MHSS for the half sample that was assigned to the SDS. See Section 7.3.5 for more details on the SDS half sample.

Originally, a weighted logistic regression model was fit on clinical diagnostic data collected from the 2008A subsample (n = 759). The dependent variable for the model was a clinical diagnosis of SMI (1 = has SMI, 0 = does not have SMI), and the predictor variables were alternative scores based on the psychological distress (K6) and function impairment (WHODAS) items collected in the NSDUH main interview. The model was used to produce a predicted probability of having SMI for each clinical interview respondent. A cut point was then established among the predicted probabilities of having SMI such that if adults with probabilities at or above the cut point were predicted to have SMI and the rest not, the weighted number of adults in the MHSS clinical sample incorrectly predicted to have SMI (false positives) would come as close as possible to equaling the weighted number of adults incorrectly predicted to not have SMI (false negatives). This approach ensures that the direct estimate of SMI based on the clinical diagnoses would be as close as possible to the predicted estimate based on the predicted SMI status for the clinical interview respondents.

Because the predictor variables in the model were variables collected during the main NSDUH interview, a probability of having SMI can be computed for every NSDUH adult respondent using the estimated model parameters. Employing the predicted probabilities from the adult NSDUH main study sample, SMI predicted values were computed for adult respondents in NSDUH main interview samples. That is, predicted values of SMI were produced (1 = predicted to have SMI; 0 = predicted not to have SMI) by applying the cut point to the individual respondent's predicted probabilities. The SMI predicted values for NSDUH

respondents then were used to compute prevalence estimates of SMI for adults. The resulting prevalence estimator is referred to as the *cut point estimator* in the remainder of this report.

The cut point determined from fitting SMI in the 2008 WHODAS clinical sample was applied to respondents in NSDUH main interview samples from 2009 and beyond to estimate SMI prevalences annually. Moreover, the probabilities of having SMI estimated from the 2008 model were used to make AMI predictions for adult NSDUH respondents. This was done by determining a second cut point such that if adults with probabilities at or above the cut point were predicted to have AMI and the rest not, the weighted total of false AMI positives and false AMI negatives in the clinical sample would come as close as possible to being equal.

Finally, a third cut point was determined so that if adults with probabilities at or above the cut point were predicted to have either serious or moderate mental illness (SMMI) and the rest not, the weighted total of false SMMI positives and false SMMI negatives in the clinical sample would come as close as possible to being equal. Predicted values for adult NSDUH respondents could then be computed for moderate mental illness (MMI = SMMI but not SMI), mild mental illness (LMI = AMI but not SMMI), and no mental illness (not AMI).

7.3 2012 Estimation Method

7.3.1 Investigating the 2008 (WHODAS) Model and Estimation Methods

By the end of 2012, a nationally representative sample of approximately 5,000 completed clinical assessments had been collected that included the 759 respondents from 2008, 520 from 2009, 516 from 2010, 1,495 from 2011, and 1,622 from 2012. This larger dataset was used to evaluate the accuracy of SMI estimates based on the 2008 estimation methods. Based on this evaluation, it was determined that revised weights and a revised model would be used to compute more accurate mental illness prevalence estimates for 2012 (see Chapter 5 for a treatment of the changes in the weights). In addition, the previously released (2008 to 2011) estimates of mental illness would be revised using the improved methods. This would not only allow for the assessment of trends in mental illness prevalence across survey years, but also provide better annual measures of mental illness across all adults and within particular subpopulations.

7.3.2 Choosing Covariates in a Model for SMI

A variety of variables was considered as predictors in a revised model of SMI used to produce cut point estimators for SMI and AMI. The following criteria were used to decide on the number and type of variables that could be included in the model. First, only a limited number of covariates could reasonably be added to a logistic model based on clinical sample data having at most, 100 effective degrees of freedom (*df*) after collapsing strata in the NSDUH full adult sample (see Section 3.7 for more information) to ensure that no primary sampling unit (PSU) is

empty (100 variance strata with two variance PSUs each). A maximum of 10 (i.e., ≈ √ 100) variables were considered for inclusion in each model.[44]

Second, some variables closely related to mental illness were avoided because including them in the model could bias the cut point estimator for a domain (i.e., subpopulation) of interest. For example, although having received services for mental health is correlated with SMI, using mental health service receipt as a covariate in a model for SMI would produce a cut point that overpredicts SMI prevalence among adults having received mental health services. Moreover, including it in the model for SMI would prevent researchers from accurately measuring changes in the receipt of treatment among those with mental illness over time.

Ultimately, the goal was to construct a parsimonious model that could be used annually to assess changes in SMI (and other categories of mental illness) within and between sociodemographic groups. A covariate related to a domain of interest would tend to fix the relationship between SMI and that domain.[45] For example, if being employed resulted in an estimated 2 percent decrease in the odds of having SMI when all other things were equal, then treating that 2 percent decrease as fixed over time would impede the measurement of any changes in the relationship between SMI and employment. Therefore, in most instances variables closely related to domain-membership indicators were not considered for inclusion into the model as covariates.

7.3.3 Metrics Used to Evaluate Models of SMI: Error Rate and Bias Measure

With the above criteria in mind, the SMI models based on the different predictor variables were evaluated using mainly two metrics: the overall error rate and domain-level bias. The error rate is a measure of the predictive power of a cut point estimator based on a particular model. It is the sum of the estimated fraction of false positives and false negatives in the adult population. Models with lower error rates produce more accurate predictions of SMI than models with higher error rates.

Different combinations of the K6 and WHODAS items and scores were evaluated as predictors in a variety of models for SMI, but none led to meaningful reductions in the error rate when compared with the alternative K6 and WHODAS scores used in the 2008 model.
By contrast, the addition of two variables from the main NSDUH interview, serious thoughts of suicide in the past year and the experience of a major depressive episode (MDE) in the past year, did noticeably decrease the error rate and were therefore to be included in a (potential) new model from SMI.

---

[44] This derives from the notion that the number of PSUs minus that number of strata must be greater than the number of estimated coefficients for the asymptotic properties of modeling fitting to be relevant. If the latter is p, the former should be at least p^2.

[45] The domains of interest were gender, age group, race/ethnicity group, region, county type, employment level, education, whether the adult received mental health treatment, whether the adult had health insurance, and the adult's household income in relation to the poverty threshold. Definitions of these domains can be found later in the text and in Table 7.3's footnotes.

The second metric used to evaluate the model was a measure of bias. A cut point estimator for SMI prevalence is based on a model, and a model is only as good as the assumptions on which it is based. As a result, a cut point estimator, unlike a model-free direct estimator computed directly from the clinical sample, can be systematically biased. A detailed description of other model-based estimators for SMI and AMI is contained in CBHSQ (in press a).

The bias measure for an SMI prevalence estimate can be measured among all adults or for domains. The bias measure for a domain was defined as the difference between the weighted proportions across the clinical sample within the domain of respondents predicted to have SMI and those actually diagnosed to have SMI (this is equal to the difference between the false positive rate and the false negative rate in the domain). Under the null hypothesis that there is no bias in the domain, this bias measure would not be significantly different from zero.

Mathematically, let y_k, c_k, w_k, and ω_k be the actual SMI diagnosis (1 = yes, 0 = no), the model-based SMI prediction, the main NSDUH weight, and the clinical sample weight, respectively, for respondent k. The cut point estimator for a domain-level prevalence is $C_D = \sum_D w_k c_k / \sum_D w_k$, where $\sum_D$ denotes summation over the domain. C_D has the same asymptotic expectation as $C_D' = \sum_D \omega_k c_k / \sum_D \omega_k$. The bias measure for C_D is B_D, the difference between C_D' and the asymptotically unbiased direct estimator $\sum_D \omega_k y_k / \sum_D \omega_k$. Note that $B_D = \sum_D \omega_k (c_k - y_k) / \sum_D \omega$ can be viewed as a simple weighted mean of $c_k - y_k$. The standard error (SE) of this mean can then be measured assuming each $c_k - y_k$ is an independent random variable with a common mean and an unknown variance that can vary across the k. Also, the ratio of B_D and its SE is asymptotically standard normal under the null hypothesis of no bias at the domain level. This ratio (and the normality assumption) was used as the t statistic in testing for the bias of a cut point estimator at the domain level.

As discussed, the cut point for SMI using the 2008 estimation method was determined so that the estimated proportion of false positives (adults predicted to have SMI but did not) and false negatives (adults predicted not to have SMI but did) in the clinical sample were as close to equal as possible. This property removed the possibility of systematic bias in the estimated proportion of adults having SMI in 2008 based on the cut point estimator of NSDUH respondents. Unfortunately, using the 2008 SMI cut point among all adults did not ensure the near equality of estimated false positives and false negatives among all adults for years other than 2008 or among domains for which SMI estimates are computed such as age groups. As a result, it was possible for the cut point estimator for a domain to be biased.

Bias in estimates of SMI was investigated for a number of domains of interest (i.e., gender, age group, race/ethnicity group, region, county type, employment level, education, whether the adult received mental health treatment, whether the adult had health insurance, and the adult's household income in relation to the poverty threshold). To do this, a model of SMI was fit on the entire WHODAS clinical sample from 2008 to 2012. Predictor variables in this model of SMI included past year K6, WHODAS, MDE, and suicidal thoughts. Results of the investigation indicated that SMI estimates within certain age groups were biased. The bias in the SMI estimates by these age groups is displayed in Table 7.1. As well as bias measures, Table 7.1

displays direct estimates of SMI[46] computed directly from the MHSS clinical sample using the clinical diagnoses and cut point estimates derived from the main NSDUH adult sample.

Table 7.1 Cut Point Estimates of SMI for Age Groups and Their Bias Measures Using a Model of SMI with Past Year K6, WHODAS, MDE, and Suicidal Thoughts as Predictor Variables: 2008A-2012 MHSS

Age	Direct Estimate[1]	Cut Point Estimate[2]	Bias Measure[3]	SE of Bias Measure[4]	*t* Value of Bias Measure[5]	*P* Value of Bias Measure[6]
All Adults	3.93	3.90	0.01	0.28	0.25	0.962
18 to 25	3.77	5.79	2.25	0.81	2.77	0.006
26 to 34	4.35	4.70	0.42	0.65	0.64	0.521
35 to 49	5.74	4.62	-1.56	0.58	-2.71	0.007
50 or Older	2.74	2.51	0.03	0.40	0.08	0.936

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SE = standard error; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

[1] Direct estimate is computed using clinical diagnoses from the clinical respondent subsample only using the weights designed for direct estimation of this sample (MHFAAWGT; see Chapter 5).

[2] Cut point estimate is computed from the main NSDUH sample of adults using predicted SMI statuses with a scaled version of the main NSDUH weights to make it comparable with the direct estimate. The scaling factors were .12 for the 2008A sample, .04 for the 2009 sample, .14 for the 2010 sample, and .35 for the 2011 and 2012 samples.

[3] Bias measure is the weighted mean value of the difference between the true value of SMI and the predicted value of SMI taken across the clinical respondent subsample within the age group (see the main text of this chapter).

[4] SE is measured as discussed in the main text of this chapter.

[5] The *t* value is the bias measure divided by its SE.

[6] The *p* value is for a two-sided test of whether the bias measure is significantly different from zero.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

As demonstrated in Table 7.1, the direct estimate is slightly larger than the cut point estimate for all adults (3.93 vs. 3.90 percent, respectively). The cut point estimate for the 18 to 25 age group, however, is much higher (5.79 percent) than the direct estimate (3.77 percent) for this age group. Moreover, the measure of the bias of this estimate is significantly different from zero at the .01 level ($p = 0.006$). These bias results, which were not paralleled in other domains of interest (see Chapter 4 of CBHSQ, in press a), suggested that adding an age-related predictor variable or variables to the model for SMI was needed to remove the bias in SMI estimates within these age groups.

It is not very plausible that, when all other factors held constant, the probability of having SMI would change suddenly when an adult aged a single year (i.e., changed from one age group to another). Consequently, a number of continuous age variables (e.g., AGE1830) were considered for addition to the SMI model. Each of these candidate variables was then compared in terms of its impact on the age group bias measures and the overall error rates for SMI, AMI, and SMMI.

The age variable that was ultimately added to the SMI model (termed AGE1830) is a recoded version of a continuous age variable for adults. It is coded as either 12 or the difference between the respondent's age and 18, whichever was smaller. The variable increased as the

[46] Mathematically, $Y_D = \sum_D \omega_k y_k / \sum_D \omega_k$.

respondent aged from 18 to 30, but then leveled off at 12 after age 30. Adding this age variable led to a cut point that both equalized false positives and false negatives for all adults and roughly within adult age groups.

After evaluating a wide variety of model specifications for both SMI and AMI that focused on finding a model that had minimum values of bias measures and error rates, the final model chosen included the following predictor variables: past year K6 score, past year WHODAS score, age, past year MDE, and past year suicidal thoughts. Table 7.2 shows how adding variables to the original 2008 WHODAS model specification, which only included the K6 and WHODAS, increased the accuracy of the SMI prevalence estimate by reducing the error rate. The estimates in this table for all adults were computed using the WHODAS sample from 2008 through 2012. The final model displayed is the 2012 (WHODAS) model.

Table 7.2 SMI Cut Point Estimates for All Adults under Different Models: 2008 to 2012 MHSS

Model Specification	Cut Point Estimate	Error Rate	False Positive Rate	False Negative Rate
K6 + WHODAS (Refit 2008 WHODAS Model)	3.97	4.42	2.21	2.21
K6 + WHODAS + AGE1830	3.89	4.30	2.17	2.12
K6 + WHODAS + MDE + SUICIDAL THOUGHTS	3.90	3.97	1.99	1.98
K6 + WHODAS + AGE1830 + MDE + SUICIDAL THOUGHTS (2012 WHODAS Model)	3.95	3.84	1.92	1.93

Age1830 = difference between respondent's age and 18 or 12, whichever was smaller; K6 = Kessler-6, a 6-item psychological distress scale; MDE = past year major depressive episode; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness; SUICIDAL THOUGHTS = suicidal thoughts in the past year.

NOTE: Scaled weights were used for both the main NSDUH sample and the clinical sample.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

7.3.4 2012 WHODAS Model

The 2012 (WHODAS) model for SMI was fit with data from 4,912 clinical interview respondents who received WHODAS in the NSDUH main interview from 2008 through 2012 with the weights designed for analyzing the entire 2008-2012 WHODAS clinical sample (MHFAAWGT, see Section 5.5 in Chapter 5). The response variable Y equaled 1 when an SMI diagnosis was positive based on the clinical interview; otherwise, Y was 0. Letting $\mathbf{X}$ be a vector of characteristics attached to a NSDUH respondent and letting the probability that this respondent had SMI be $\pi = \Pr(Y = 1 \mid \mathbf{X})$, the 2012 model was

$$\begin{aligned}\text{logit}(\hat{\pi}) = \log[\hat{\pi} / (1 - \hat{\pi})] = -5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s \\ + 1.1267330\,X_m + 0.1059137X_a\end{aligned} \tag{7.1}$$

or

$$\hat{\pi} = \frac{1}{1 + \exp[-(-5.972664 + 0.0873416X_k + 0.3385193X_w + 1.9552664X_s + 1.1267330\,X_m + 0.1059137X_a)]},$$

where $\hat{\pi}$ is the estimated probability an adult had SMI. The covariates in equation (7.1) come from the main NSDUH interview data:

- X_k = *Alternative Past Year K6 Score* (WSPDSC2[47]): Past year K6 score of less than 8 recoded as 0; past year K6 score of 8 to 24 recoded as 1 to 17.
- X_w = *Alternative WHODAS Score* (WHODASC3): WHODAS item score of less than 2 recoded as 0; WHODAS item score of 2 to 3 recoded as 1, then summed for a score ranging from 0 to 8.
- X_s = *Serious Thoughts of Suicide in the Past Year* (MHSUTK_U): Coded as 1 if "yes"; coded as 0 otherwise.
- X_m = *Past Year MDE* (AMDEY2_U): Coded as 1 if the criteria for past year MDE were met; coded as 0 otherwise.
- X_a = *AGE1830* (Recoded Age): Coded as age minus 18 if aged 18 to 30; coded as 12 otherwise.

A cut point probability π_0 was determined, so that if $\hat{\pi} \geq \pi_0$ for a particular respondent, then he or she was predicted to be SMI positive; otherwise, he or she was predicted to be SMI negative. The cut point (0.260573529) was chosen so that the weighted numbers of false positives and false negatives in the MHSS dataset were as close to equal as possible. The predicted SMI status for all adult NSDUH respondents was used to compute prevalence estimates of SMI.

A second cut point probability (0.0192519810) was determined so that any respondent with an SMI probability greater than or equal to the cut point was predicted to be positive for AMI, and the remainder were predicted to be negative for AMI. The second cut point was chosen so that the weighted numbers of AMI false positives and false negatives were as close to equal as possible.

Estimates of SMMI (GAF score below 60) were analogously computed with the SMI method; the cut point was 0.077686285365. Estimates of LMI and MMI were derived by a process of subtraction. Respondents were classified as belonging to the MMI category if they belonged to the SMMI category, but they did not belong to the SMI category. Respondents were classified as belonging to the LMI category if they belonged to the AMI category, but not to the SMMI category.

7.3.5 2012 Model for the SDS Sample

One of the predictors in the original 2008 SDS model for SMI was a variable based on the SDS, similar to the WHODAS variable in the WHODAS-based model. This model was selected so that it provided SMI estimates that were comparable with the 2008 WHODAS model-based SMI estimates. However, when that model was applied to the 2008B MHSS clinical sample to produce an AMI prevalence estimate for all adults, the result (16.73 percent) was not close enough to the AMI estimates for 2008A and 2009 calculated using the 2008

[47] See Section 4.4 in Chapter 4 for information on how the missing values were treated for the predictor variables in the 2012 model.

WHODAS model (19.48 and 19.86 percent, respectively). See Aldworth et al. (2010, pp. 36-38). As a result, AMI (and SMMI) prevalence estimates derived from the 2008B sample were not previously reported, for example, in the 2011 NSDUH mental health findings report (CBHSQ, 2012c).

This led to a search for a better model for the 2008 SDS sample so that the data from this sample could be used when producing estimates of SMI and AMI in 2008. The resulting 2012 model for SDS had the same covariates as the 2012 WHODAS model except that the WHODAS variable was removed. In addition, for the 2008B respondents, an adjusted version of MDE variable was used.[48] The model was fit to the entire clinical sample from 2008 through 2012 with the weights designed for analyzing the entire 2008-2012 clinical sample (MHFNLWGT, see Section 5.5 in Chapter 5). This model was then used to predict mental illness statuses for the 2008B main NSDUH sample.

In order to determine whether the predicted values from the 2008B sample could be combined with predicted values from the complete WHODAS sample when making mental illness prevalence estimates, an investigation was conducted to determine whether the estimates produced using this 2012 SDS model differed from those using the 2012 WHODAS model. Also investigated was whether the estimates produced using this 2012 SDS model significantly differed from direct estimates. Tables 7.3 and 7.4 show small differences between domain and all-adult estimates for 2008 SMI and AMI derived from the 2008A NSDUH respondent sample using the 2012 WHODAS model and the 2008B NSDUH respondent sample using the 2012 SDS model. The tables also contain the statistics on the bias measure for the 2008B sample estimates. No bias measure is significantly different from zero at the 0.1 level. Consequently, it was determined that the predicted values from the 2008B sample could be combined with predicted values from the complete WHODAS sample for 2008A and for 2009 through 2012 when making mental illness prevalence estimates.

The 2012 SMI prediction model for the SDS sample is

$$\text{logit}(\hat{\pi}) = \log[\hat{\pi} / (1 - \hat{\pi})] = -5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a \qquad (7.2)$$

or

$$\hat{\pi} = \frac{1}{1 + \exp[-(-5.7736246 + 0.1772067X_k + 1.8392433X_s + 1.6428623X_m + 0.1231266X_a)}.$$

All of the covariates in equation (7.2) appeared in equation (7.1) as well.

[48] The variable was developed to remove the context effects on MDE in the 2008B NSDUH instrument. The MDE data for respondents in the SDS sample were not comparable with the MDE data for the respondents in the WHODAS sample. This was because respondents differentially answered the MDE questions (i.e., context effects) based on whether they had answered the WHODAS or SDS item in the preceding questionnaire module. For more details, see Aldworth et al. (2012).

Table 7.3 Comparing Cut Point Estimates for SMI from the 2008B Sample Using the 2012 Model for the SDS Sample with Estimates from the 2008A Sample Using the 2012 Model for the WHODAS Sample

	Cut Point Estimates			Bias Measures for 2008B Estimates		
Domain	**2008A**	**2008B**	**Difference**	**Bias Measure**	**Standard Error of Bias Measure**	***P* Value of Bias Measure**
All Adults	3.72	3.70	-0.02	-1.43	1.91	0.455
Age: 18 to 25	4.15	3.42	-0.73	-0.77	0.83	0.355
Age: 26 to 34	4.57	4.83	0.26	-0.48	2.16	0.825
Age: 35 to 49	4.74	4.92	0.18	2.11	2.74	0.442
Age: 50 or Older	2.52	2.52	0.00	-4.37	4.17	0.294
Health Insurance: Yes	3.54	3.44	-0.10	-0.68	2.02	0.736
Health Insurance: No	4.71	5.17	0.46	-9.14	5.79	0.115
< 100% of the Poverty Threshold[1]	6.25	5.61	-0.64	-16.80	11.29	0.136
100%-199% of the Poverty Threshold[1]	4.84	5.11	0.27	0.15	1.31	0.912
≥ 200% of the Poverty Threshold[1]	3.02	3.02	0.00	1.21	1.16	0.296
White, Not Hispanic	4.14	4.27	0.13	-2.46	2.47	0.319
Black, Not Hispanic	2.66	2.31	-0.35	-1.01	4.80	0.833
Other, Not Hispanic	3.40	2.46	-0.94	5.19	10.18	0.610
Hispanic	2.58	2.56	-0.02	0.31	0.38	0.418
Male	2.76	2.50	-0.26	-0.34	1.74	0.846
Female	4.61	4.83	0.22	-2.45	3.29	0.457
Northeast	3.36	4.35	0.99	-4.96	5.49	0.367
Midwest	4.23	3.75	-0.48	-0.02	0.56	0.970
South	3.71	3.14	-0.57	-2.46	2.09	0.239
West	3.53	4.02	0.49	4.82	4.41	0.275
Large Metro	3.55	3.60	0.05	1.71	1.71	0.316
Small Metro	3.82	4.15	0.33	-3.77	4.60	0.413
Nonmetro	4.07	3.19	-0.88	-3.84	2.86	0.179
Received Mental Health Treatment: Yes[2]	18.39	17.90	-0.49	-16.70	11.23	0.137
Received Mental Health Treatment: No[2]	1.43	1.50	0.07	0.94	0.91	0.301
Employed Full Time	2.69	3.09	0.40	0.95	1.68	0.571
Employed Part Time	4.06	3.40	-0.66	0.37	1.09	0.735
Unemployed	7.98	5.70	-2.28	0.98	2.31	0.672
Other Employment Status[3]	5.00	4.70	-0.30	-6.01	5.07	0.235
Less than High School	3.28	3.47	0.19	-10.30	7.07	0.144
High School Graduate	4.38	3.46	-0.92	0.25	0.53	0.635
Some College	4.33	4.88	0.55	-3.78	5.52	0.494
College Graduate	2.65	3.04	0.39	1.53	2.61	0.558

NSDUH = National Survey on Drug Use and Health; SDS = Sheehan Disability Scale; SMI = serious mental illness; WHODAS = World Health Organization Disability Assessment Schedule.

[1] Poverty Threshold incorporates information on family income, size, and composition and is calculated as a percentage of the U.S. Census Bureau's poverty thresholds. Respondents aged 18 to 22 who were living in a college dormitory were excluded.

[2] Received Mental Health Treatment is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.

[3] Other Employment includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

Table 7.4 Comparing Cut Point Estimates for AMI from 2008B Sample Using the 2012 Model for the SDS Sample with Estimates from the 2008A Sample Using the 2012 Model for the WHODAS Sample

	Cut Point Estimates			Bias Measures for 2008B Estimates		
Domain	**2008A**	**2008B**	**Difference**	**Bias Measure**	**Standard Error of Bias Measure**	***P* Value of Bias Measure**
All Adults	17.69	17.78	0.09	-1.58	2.75	0.566
Age: 18 to 25	18.83	18.44	-0.39	-2.17	3.48	0.532
Age: 26 to 34	20.75	23.14	2.39	-11.20	7.77	0.150
Age: 35 to 49	20.08	20.19	0.11	4.17	5.02	0.405
Age: 50 or Older	14.43	13.79	-0.64	-0.93	4.57	0.839
Health Insurance: Yes	17.00	16.62	-0.38	-1.18	2.97	0.692
Health Insurance: No	21.43	24.44	3.01	-5.76	5.67	0.310
< 100% of the Poverty Threshold[1]	25.01	25.72	0.71	-18.90	11.51	0.100
100%-199% of the Poverty Threshold[1]	22.39	21.15	-1.24	6.02	4.43	0.174
≥ 200% of the Poverty Threshold[1]	15.27	15.60	0.33	0.07	2.87	0.981
White, Not Hispanic	18.44	18.10	-0.34	-3.16	3.44	0.359
Black, Not Hispanic	15.79	17.60	1.81	-10.60	6.56	0.105
Other, Not Hispanic	16.70	16.89	0.19	17.86	9.85	0.070
Hispanic	15.91	16.70	0.79	4.75	7.42	0.522
Male	13.45	13.80	0.35	0.84	3.84	0.827
Female	21.64	21.50	-0.14	-3.84	3.90	0.324
Northeast	16.68	21.16	4.48	-6.49	7.11	0.361
Midwest	18.64	18.04	-0.60	4.97	2.71	0.067
South	17.50	16.49	-1.01	-6.67	3.82	0.081
West	17.89	16.86	-1.03	7.35	7.28	0.313
Large Metro	16.91	17.57	0.66	-1.09	4.00	0.785
Small Metro	18.14	18.61	0.47	-3.72	5.69	0.514
Nonmetro	19.35	16.93	-2.42	1.00	2.99	0.738
Received Mental Health Treatment: Yes[2]	53.92	53.85	-0.07	-2.19	14.34	0.879
Received Mental Health Treatment: No[2]	12.05	12.16	0.11	-1.49	2.26	0.511
Employed Full Time	15.01	16.38	1.37	-0.97	3.83	0.800
Employed Part Time	20.86	16.96	-3.90	0.33	4.15	0.936
Unemployed	23.72	25.45	1.73	0.74	13.94	0.958
Other Employment Status[3]	20.55	19.66	-0.89	-3.49	5.48	0.524
Less than High School	19.09	21.57	2.48	-0.68	6.17	0.912
High School Graduate	17.77	16.31	-1.46	-0.11	2.43	0.964
Some College	18.55	20.13	1.58	-6.82	7.10	0.337
College Graduate	16.04	15.13	-0.91	2.07	5.33	0.698

AMI = any mental illness; NSDUH = National Survey on Drug Use and Health; SDS = Sheehan Disability Scale; WHODAS = World Health Organization Disability Assessment Schedule.

[1] Poverty Threshold incorporates information on family income, size, and composition and is calculated as a percentage of the U.S. Census Bureau's poverty thresholds. Respondents aged 18 to 22 who were living in a college dormitory were excluded.

[2] Received Mental Health Treatment is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.

[3] Other Employment includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

7.4 Comparisons of SMI and AMI Estimates Using the 2012 and 2008 Estimation Methods

Chapter 5 and the previous sections of this chapter discuss why and how the methods of weighting the clinical sample data and then modeling that data to produce SMI and AMI estimates have changed. This section investigates the effects of changes in estimation methodology on the SMI and AMI estimates. Tables 7.5 and 7.6 display SMI and AMI prevalence estimates for the combined time period 2008-2012 computed using the opposing estimation methods, that is, the NSDUH adults mental illness estimates using the 2008 model and weights versus the 2012 model and weights.

The SMI estimates for all adults in the NSDUH sample in these two tables for 2008-2012 have been computed using NSDUH analysis weights. In this combined prevalence estimate, the estimate from 2012 "counted" slightly more than the estimate from 2008 because the adult population size in 2012 was slightly larger. The estimates from the 2008 WHODAS sample (the 2008A sample) and SDS (2008B) sample were each implicitly weighted by half the number of adults in 2008 with one exception. Only the 2008A sample was used for AMI estimates with the 2008 method, with the estimates of that half sample then implicitly scaled by the adult population in 2008.

By contrast, the direct estimates in the tables were computed with the final scaled clinical sample weights as described in Section 5.5. The final clinical sample weights were scaled for use in analyses with combined years of data to increase statistical efficiency.[49] This scaling has the potential of biasing the resulting estimates. An *F* test applied to the all-adult estimates based on the clinical sample revealed no significant differences among the years in either SMI or AMI prevalences at the 0.1 level.

Table 7.5 presents the estimates of SMI for the overall population and within several domains of interest. As shown in the table, the 2012-method-based SMI estimates for the adult population and within domains tended to be smaller than their corresponding 2008-method-based counterparts. This is due mostly to the changes in the weights used in fitting the models. Recall that there should be almost no bias in the standard cut point estimates for all adults due to model fitting because of attempts to equalize the weighted false positive and false negative rates. Nevertheless, the 2008 method-based estimate has a significant upward bias measure, which must be the result of it being based on a fit using different weights.

Most of the 2012 method-based estimates in Table 7.5 were closer to the nearly unbiased direct estimates from the clinical sample than the 2008-method-based estimates. This is especially true for the 18 to 25 age group where the estimate using the 2008 method has a large bias. Only one domain in Table 7.5, the poverty level of less than a 100 percent threshold, had a significantly biased SMI estimate when applying the 2012 method. That is not wholly unexpected under the null hypothesis of no domain-level biases, given the number of tests conducted and therefore type I error inflation (64).

[49] The scaling factors were .06 for the 2008A and 2008B clinical samples, .04 for the 2009 sample, .14 for the 2010 sample, and .35 for the 2011 and 2012 samples.

Table 7.5 Comparing SMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample

Domain	Direct Estimate from Clinical Sample	Estimate Using 2008 Models	Estimate Using 2012 Model
Overall	3.9	4.9[a]	3.9
Age			
18 to 25	3.7	7.7[a]	3.8
26 to 34	4.2	6.5[a]	5.0
35 to 49	5.7	5.5	5.0
50 or Older	2.8	3.0	2.8
Gender			
Male	3.0	3.4	2.9
Female	4.9	6.3[a]	4.8
Race			
White, Not Hispanic	4.4	5.2	4.3
Black, Not Hispanic	3.5	3.9	3.1
Other, Not Hispanic	4.5	4.4	3.1
Hispanic	2.0	4.4	3.1
Region			
Northeast	3.1	4.5	3.7
Midwest	4.0	5.1[a]	4.2
South	3.9	4.8	3.7
West	4.8	5.2	4.1
County Type			
Large Metro	3.7	4.6	3.6
Small Metro	4.2	5.2[a]	4.2
Nonmetro	4.2	5.3	4.2
Received Mental Health Treatment[1]			
Yes	19.4	21.3	18.5
No	1.5	2.3[a]	1.6
Employment Level			
Employed Full Time	2.4	3.6[a]	2.8
Employed Part Time	4.2	5.5[a]	4.0
Unemployed	5.3	7.8	6.3
Other[2]	6.2	6.3	5.2
Education			
Less than High School	5.8	5.5	4.0
High School Graduate	3.8	5.1	4.0
Some College	4.5	5.9[a]	4.5
College Graduate	2.9	3.5	3.1

See notes at end of table. (continued)

Table 7.5 Comparing SMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample (continued)

Domain	Direct Estimate from Clinical Sample	Estimate Using 2008 Models	Estimate Using 2012 Model
Poverty Threshold[3]			
< 100% Threshold	9.9	9.1	6.9[a]
100% - 199% Threshold	5.3	6.1[a]	4.8
≥ 200% Threshold	2.5	3.7[a]	3.0
Health Insurance			
Yes	3.6	4.6[a]	3.6
No	5.8	6.7	5.2

NSDUH = National Survey on Drug Use and Health; SMI = serious mental illness.

NOTE: The weights were *not* scaled for the NSDUH estimators when combining data across years. For the direct estimator, the scaling factors were .06 for the 2008A and 2008B clinical samples, .04 for the 2009 sample, .14 for the 2010 sample, and .35 for the 2011 and 2012 samples.

[a] The difference between the model-based and direct estimates computed from the subclinical sample is significantly different from zero at the .05 level.

[1] Received Mental Health Treatment is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.

[2] Other Employment includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

[3] Poverty Threshold incorporates information on family income, size, and composition and is calculated as a percentage of the U.S. Census Bureau's poverty thresholds. Respondents aged 18 to 22 who were living in a college dormitory were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

Table 7.6 presents the estimates of AMI for the overall population and within the same domains of interest (AMI estimates computed with the 2008 model do not use data from the 2008B sample). Similar to the SMI results, almost all of the 2012 method-based AMI estimates were smaller than their corresponding 2008 method-based estimates, except for persons aged 50 or older. As with SMI, the 18 to 25 age group had a very large bias measure using the 2008 method. Moreover, most of the 2012 method-based estimates were closer to the direct estimates from the MHSS clinical sample than the 2008 method-based estimates, particularly the AMI estimates for the younger age groups. One subpopulation, the South region, had a significantly biased AMI estimate when applying the 2012 model. However, both the direct estimate for the South and the cut point estimate based on the 2012 model were below the national average.

Table 7.6 Comparing AMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample

Domain	Direct Estimate from Clinical Sample	Estimate Using 2008 Model	Estimate Using 2012 Model
Overall	18.0	20.0	18.1
Age			
18 to 25	20.9	30.5[a]	18.5
26 to 34	19.7	24.1	22.0
35 to 49	20.2	20.5	20.4
50 or Older	14.9	14.4	14.9
Gender			
Male	14.4	16.1	14.4
Female	21.3	23.5	21.5
Race			
White, Not Hispanic	18.2	20.5	19.0
Black, Not Hispanic	15.8	19.4	16.8
Other, Not Hispanic	16.3	19.3	16.8
Hispanic	19.4	17.9	15.3
Region			
Northeast	19.4	19.6	18.0
Midwest	16.4	20.3	18.3
South	16.9	19.7[a]	17.7[a]
West	20.0	20.4	18.4
County Type			
Large Metro	19.3	19.2	17.4
Small Metro	16.9	20.7	18.9
Nonmetro	15.8	20.9[a]	18.6
Received Mental Health Treatment[1]			
Yes	53.2	55.4	53.9
No	12.4	14.3	12.3
Employment Level			
Employed Full Time	15.0	16.6	15.4
Employed Part Time	19.7	23.8	19.3
Unemployed	20.7	28.6	23.8
Other[2]	21.8	22.2[a]	21.0
Education			
Less than High School	25.9	22.5	19.9
High School Graduate	17.2	19.8	17.7
Some College	16.4	21.9[a]	19.6
College Graduate	16.7	17.0	16.1

See notes at end of table. (continued)

Table 7.6 Comparing AMI Prevalence Estimates Using the 2008 and 2012 Models: 2008-2012 Adult NSDUH Main Study and Clinical Sample (continued)

Domain	Direct Estimate from Clinical Sample	Estimate Using 2008 Model	Estimate Using 2012 Model
Poverty Threshold[3]			
< 100% Threshold	25.5	29.9	25.6
100% - 199% Threshold	24.3	23.6	20.9
≥ 200% Threshold	14.9	16.8	15.7
Health Insurance			
Yes	16.9	19.0	17.4
No	23.4	24.8	21.4

AMI = any mental illness; NSDUH = National Survey on Drug Use and Health.

NOTE: The weights were *not* scaled for the cut point estimators when combining data across years. For the direct estimator, the scaling factors were .06 for the 2008A and 2008B clinical samples, .04 for the 2009 sample, .14 for the 2010 sample, and .35 for the 2011 and 2012 samples.

[a] The difference between the model-based and direct estimates computed from the subclinical sample is significantly different from zero at the .05 level.

[1] Received Mental Health Treatment is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.

[2] Other Employment includes students, persons keeping house or caring for children full time, retired or disabled persons, or other persons not in the labor force.

[3] Poverty Threshold incorporates information on family income, size, and composition and is calculated as a percentage of the U.S. Census Bureau's poverty thresholds. Respondents aged 18 to 22 who were living in a college dormitory were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality; NSDUH main study and clinical sample, 2008-2012.

7.5 Caveats on Model-Based Mental Illness Prevalence Estimates

Various caveats on the methods from producing model-based estimates of SMI should be noted. NSDUH covers only residents of households (i.e., persons living in houses/townhouses, apartments, condominiums; civilians living in housing on military bases, etc.) and persons in noninstitutional group quarters (e.g., shelters, rooming/boarding houses, college dormitories, migratory workers' camps, halfway houses). It does not cover persons who, for the entire year, had no fixed address (e.g., homeless and/or transient persons not in shelters), were on active military duty, or who resided in institutional group quarters (e.g., correctional facilities, nursing homes, mental institutions, and long-term hospitals). Many persons in these excluded categories have mental illness (especially homeless persons and those living in institutional group quarters), but they are not accounted for in the NSDUH SMI and AMI prevalence estimates.

Although effort went into adjusting for the potential biases in those adults responding to the clinical interview, there is no guarantee that the adjustments for nonresponse to the clinical and the undercoverage of Hispanics who chose to respond to the main survey in Spanish were completely successful. There is likewise no guarantee that the nonresponse and coverage adjustments in NSDUH were completely successful in removing all of the biases.

In addition, the mental illness estimates were based on a weighted logistic model for SMI. Although statistical tests did not uncover significant biases in the SMI prevalence estimates within most of the key subpopulations of interest, that does not guarantee that other subpopulation-level biases do not exist. Although many of the subpopulation-level estimates commonly produced were investigated, there are many more possible subpopulations for which SMI estimates may be computed given the vast number of variables available in the NSDUH data. Also, it should be noted that SMI prevalence estimates for a subpopulation closely related to a variable used in the SMI model is likely to be biased. Such variables include suicidal thoughts, the experience of MDE, and the various components of the K6 or WHODAS scales. An ongoing research study is evaluating several alternate models that do not use suicidal thoughts and the experience of MDE as predictors and may produce unbiased estimates for the corresponding subpopulations.

Another caveat concerns the SEs of the SMI estimates. By treating model-predicted SMI indicators as true values of SMI when estimating the SE of a cut point estimate, SEs are underestimated. This underestimation is due to not accounting for the error in model fitting. Nevertheless, SEs calculated in this way can be useful when estimating the difference in SMI prevalence between subpopulations or differences in estimates over time because the same model fit is used for both subpopulations; therefore, the error due to model fitting when estimating differences is effectively cancelled out.

Finally, the MHSS clinical data collection ended in 2012, and the 2012 SMI model will be applied to NSDUH data going forward to produce mental illness estimates under the assumption that the relationship between SMI and the predictor variables is stable over time. Without continuing the clinical interview, the validity of this assumption cannot be assessed, and one cannot test whether covariates other than those specified in the 2012 model may be needed in the future to reduce misclassification or subpopulation-level biases.

References

Aldworth, J., Barnett-Walker, K., Chromy, J., Karg, R., Kott, P., & Morton, K. (2010, December). *Measuring serious mental illness with the NSDUH: Results of 2009 12-month analysis* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022, Mental Health Surveillance Survey Deliverable No. 7, RTI/0209009.523.006.008). Research Triangle Park, NC: RTI International.

Aldworth, J., Barnett-Walker, K., Chromy, J., Karg, R., Morton, K., Novak, S., & Spagnola, K. (2009, June). *Measuring serious mental illness with the NSDUH: Results of 2008 12-month analysis* (prepared for the Substance Abuse and Mental Health Services Administration under Contract No. 283-2004-00022, Mental Health Surveillance Survey Deliverable No. 5, RTI/0209009.423.006.008). Research Triangle Park, NC: RTI International.

Aldworth, J., Kott, P., Yu, F., Mosquin, P., & Barnett-Walker, K. (2012). Analysis of effects of 2008 NSDUH questionnaire changes: Methods to adjust adult MDE and SPD estimates and to estimate SMI in the 2005-2009 surveys. In *2010 National Survey on Drug Use and Health: Methodological resource book* (Section 16b, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283200800004C, Deliverable No. 39, RTI/0211838.108.005). Research Triangle Park, NC: RTI International.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (DSM-IV) (4th ed.). Washington, DC: Author.

American Psychiatric Publishing, Inc. (2000). *Diagnostic and statistical manual of mental disorders, 4th ed., text revision* (DSM-IV-TR; doi: 10.1176/appi.books.9780890423349). Retrieved from http://www.psychiatryonline.com/resourceTOC.aspx?resourceID=1

Center for Behavioral Health Statistics and Quality. (2010). *Results from the 2009 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 10-4609, NSDUH Series H-39). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2011). *Results from the 2010 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 11-4658, NSDUH Series H-41). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2012a). *Results from the 2010 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 11-4667, NSDUH Series H-42). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2012b). *Results from the 2011 National Survey on Drug Use and Health: Summary of national findings* (HHS Publication No. SMA 12-4713, NSDUH Series H-44). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (2012c). *Results from the 2011 National Survey on Drug Use and Health: Mental health findings* (HHS Publication No. SMA 12-4725, NSDUH Series H-45). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2013). *National Survey on Drug Use and Health: 2012 public use file and codebook* (doi:10.3886/ICPSR34933.v1). Retrieved from http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/34933

Center for Behavioral Health Statistics and Quality. (2014a). Mental Health Surveillance Study operations report (2008-2012). In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 16a). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration. (2014b). *National Survey on Drug Use and Health: 2008-2012 adult clinical interview data file and codebook*. Retrieved from http://www.icpsr.umich.edu/icpsrweb/content/SAMHDA/dataportal.html

Center for Behavioral Health Statistics and Quality. (in press a). *Estimating mental illness among adults in the United States: Revisions to 2008 estimation procedures*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Center for Behavioral Health Statistics and Quality. (in press b). *Evaluation of imputation methods for the NSDUH*. Rockville, MD: Substance Abuse and Mental Health Services Administration.

Chen, P., Cribb, D., Dai, L., Gordek, H., Laufenberg, J., Sathe, N., & Westlake, M. (2014). Person-level sampling weight calibration. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C, Phase I, Deliverable No. 41, RTI/0212800.001.107.004). Research Triangle Park, NC: RTI International.

Clark, W. B., & Hill, M. H. (Eds.). (1991). *Alcohol in America: Drinking practices and problems*. Albany, NY: State University of New York Press.

Colpe, L. J., Barker, P. R., Karg, R. S., Batts, K. R., Morton, K. B., Gfroerer, J. C., Stolzenberg, S. J., Cunningham, D. B., First, M. B., & Aldworth, J. (2010). The National Survey on Drug Use and Health Mental Health Surveillance Study: Calibration study design and field procedures. *International Journal of Methods in Psychiatric Research, 19*(Suppl. 1), 36-48. doi: 10.1002/mpr.311

Crippa, J. A., de Lima Osório, F., Del-Ben, C. M., Filho, A. S., da Silva Freitas, M. C., & Loureiro, S. R. (2008). Comparability between telephone and face-to-face structured clinical interview for DSM-IV in assessing social anxiety disorder. *Perspectives in Psychiatric Care, 44*, 241-247.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*, 766-771.

Fennig, S., Craig, T., Lavelle, J., Kovasznay, B., & Bromet, E. J. (1994). Best-estimate versus structured interview-based diagnosis in first-admission psychosis. *Comprehensive Psychiatry, 35,* 341-348.

First, M. B., Spitzer R. L., Gibbon M., & Williams J. B. W. (2002). *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition. (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute, Biometrics Research Department.

Furukawa, T. A., Kessler, R. C., Slade, T., & Andrews, G. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine, 33*, 357-362.

Greenfield, T. K., & Kaskutas, L. A. (1998). Five years' exposure to alcohol warning label messages and their impacts: Evidence from diffusion analysis. *Applied Behavioral Science Review, 6*, 39-68.

Hajebi, A., Motevalian, A., Amin-Esmaeili, M., Hefazi, M., Radgoodarzi, R., Rahimi-Movaghar, A., & Sharifi, V. (2012). Telephone versus face-to-face administration of the Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, for diagnosis of psychotic disorders. *Comprehensive Psychiatry, 53*, 579-583.

Jackson, J. S., Neighbors, H. W., Nesse, R. M., Trierweller, S. J., & Torres, M. (2004). Methodological innovations in the National Survey of American Life. *International Journal of Methods in Psychiatric Research, 13*, 289-298.

Jordan, B. K., Karg, R. S., Batts, K. R., Epstein, J. F., & Wiesen, C. A. (2008). A clinical validation of the National Survey on Drug Use and Health Assessment of Substance Use Disorders. *Addictive Behaviors, 33*, 782-798.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., & Eaves, L.J. (1992). A population based twin study of major depression in women: The impact of varying definitions of illness. *Archives of General Psychiatry, 49*, 257-266.

Kessler, R., Abelson, J., Demler, O., Escobar, J. I., Gibbon, M., Guyer, M. E., Howes, M. J., Jin, R., Vega, W. A., Walters, E. E., Wang, P., Zaslavsky, A., & Zheng, H. (2004). Clinical calibration of DSM-IV diagnoses in the World Mental Health (WMH) version of the World Health Organization (WHO) Composite International Diagnostic Interview (WMH-CIDI). *International Journal of Methods in Psychiatric Research, 13*(2), 122-139.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., Howes, M. J., Normand, S. L., Manderscheid, R. W., Walters, E. E., & Zaslavsky, A. M. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*, 184-189. doi: yoa20567 [pii]

Kott, P. (2011). A nearly pseudo-optimal method for keeping calibration weights from falling below unity in the absence of nonresponse or frame errors. *Pakistan Journal of Statistics, 27*, 391-396.

Kranzler, H. R., Kadden, R. M., Babor, T. F., Tennen, H., & Rounsaville, B. J. (1996). Validity of the SCID in substance abuse patients. *Addiction, 91,* 859-868.

Kranzler, H. R., Kadden, R. M., Burleson, J. A., Babor, T. F., Apter, A., & Rounsaville, B. J. (1995). Validity of psychiatric diagnoses in patients with substance use disorders: Is the interview more important than the interviewer? *Comprehensive Psychiatry, 36*, 278-288.

Lee, S., Tsang, A., Lau, L., Mak, A., Ng, K. L., & Chan, D. M. (2008). Concordance between telephone survey classification and face-to-face structured clinical interview in the diagnosis of generalized anxiety disorder in Hong Kong. *Journal of Anxiety Disorders, 22*, 1403-1411.

Leon, A. C., Olfson, M., Portera, L., Farber, L., & Sheehan, D. V. (1997). Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. *International Journal of Psychiatry in Medicine, 27*(2), 93-105.

Liao, D., Aldworth, J., Yu, F., Morton, K., Chen, P., Shook-Sa, B., Kott, P., Davis, T., & Karg, R. (2012). *2011 Mental Health Surveillance Study: Design and estimation report* (prepared for the Substance Abuse and Mental Health Services Administration, Contract No. HHSS283200800004C, RTI 0211838.212.008). Research Triangle Park, NC: RTI International.

Lohr, S. L. (1999). *Sampling: Design and analysis.* Belmont, CA: Duxbury Press.

Maddahian, E., Newcomb, M. D., & Bentler, P. M. (1988). Adolescent drug use and intention to use drugs: Concurrent and longitudinal analyses of four ethnic groups. *Addictive Behaviors, 13*, 191-195.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2012, June). Sample design report. In *2011 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration, under Contract No. HHSS283200800004C, Phase II, Deliverable No. 8, RTI/0211838.203.004). Research Triangle Park, NC: RTI International.

Morton, K. B., Martin, P. C., Shook-Sa, B. E., Chromy, J. R., & Hirsch, E. L. (2013, January). Sample design report. In *2012 National Survey on Drug Use and Health: Methodological resource book* (Section 2, prepared for the Substance Abuse and Mental Health Services Administration under Contract No. HHSS283201000003C, Phase I, Deliverable No. 8, RTI/0212800.001.103). Research Triangle Park, NC: RTI International.

Novak, S. P., Colpe, L. J., Barker, P. R., & Gfroerer, J. C. (2010). Development of a brief mental health impairment scale using a nationally representative sample in the USA. *International Journal of Methods in Psychiatric Research*, *19*(Suppl. 1), 49-60. doi:10.1002/mpr.313

Office of Applied Studies. (2009). *Results from the 2008 National Survey on Drug Use and Health: National finding*s (HHS Publication No. SMA 09-4434, NSDUH Series H-36). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010a). *Results from the 2009 National Survey on Drug Use and Health: Volume I. Summary of national findings* (HHS Publication No. SMA 10-4586Findings, NSDUH Series H-38A). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Office of Applied Studies. (2010b). *Results from the 2009 National Survey on Drug Use and Health: Volume II. Technical appendices and selected prevalence tables* (HHS Publication No. SMA 10-4586Appendices, NSDUH Series H-38B). Rockville, MD: Substance Abuse and Mental Health Services Administration.

Ramirez Basco, M., Bostic, J. Q., Davies, D., Rush, A. J., Witte, B., Hendrickse, W., & Barnett, V. (2000). Methods to improve diagnostic accuracy in a community mental health setting. *American Journal of Psychiatry, 157*, 1599-1605.

Rehm, J., Üstün, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., & Adlaf, E. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*(2), 110-123. doi:10.1002/mpr.61

Rohde, P., Lewinsohn, P.M., & Seeley, J.R. (1997). Comparability of telephone and face-to-face interviews in assessing axis I and axis II disorders. *American Journal of Psychiatry, 154*, 1593-1598.

Ross, M., Xun, W. Q. E., & Wilson, A. (2002). Language and the bicultural self. *Personality and Social Psychology Bulletin, 28*, 1040-1050.

RTI International. (2012). *SUDAAN®, Release 11.0* [computer software]. Research Triangle Park, NC: Author.

Segal, D. L., Kabacoff, R. I., Hersen, M., Van Hasselt, V. B., & Ryan, C. F. (1995). Update on the reliability of diagnosis in older psychiatric outpatients using the Structured Clinical Interview for DSM-III-R. *Journal of Clinical Geropsychology, 1,* 313-321.

Shear, M. K., Greeno, C., Kang, J., Ludewig, D., Frank, E., Swartz, H. A., & Hanekamp, M. (2000). Diagnosis of nonpsychotic patients in community clinics. *American Journal of Psychiatry, 157*, 581-587.

Silva de Crane, R., & Spielberger, C. D. (1981). Attitudes of Hispanic, black and Caucasian university students toward mental illness. *Hispanic Journal of Behavioral Sciences, 3*, 241-255.

Sobin, C., Weissman, M. M., Goldstein, R. B., Adams, P., Wickramaratne, P., Warner, V., & Lish, J. D. (1993). Diagnostic interviewing for family studies: Comparing telephone and face-to-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatric Genetics, 3*, 227-233.

Steiner, J. L., Tebes, J. K., Sledge, W. H., & Walker, M. L. (1995). A comparison of the structured clinical interview for DSM-III-R and clinical diagnoses. *Journal of Nervous and Mental Disease, 183,* 365-369.

Substance Abuse and Mental Health Services Administration, Center for Mental Health Services. (1993, May 20). Final notice [Final definitions for: (1) Children with a serious emotional disturbance, and (2) adults with a serious mental illness]. *Federal Register, 58*(96), 29422-29425.

Trafimow, D., Silverman, E. S., Mei-Tai Fan, R., & Shui Fun Law, J. (1997). The effect of language and priming on the relative accessibility of the private self and the collective self. *Journal of Cross-Cultural Psychology, 28*, 107-123.

Triandis, H. C., Davis, E. E., Vassiliou, V., & Nassiakou, M. (1965). *Some methodological problems concerning research on negotiations between monolinguals*. Urbana, IL: University of Illinois.

Zanarini, M. C., & Frankenburg, F. R. (2001). Attainment and maintenance of reliability of axis I and II disorders over the course of a longitudinal study. *Comprehensive Psychiatry, 42,* 369-374.

Zanarini, M. C, Skodol, A. E., Bender, D., Dolan, R., Sanislow, C., Schaefer, E., Morey, L. C., Grilo, C. M., Shea, M. T., McGlashan, T. H., & Gunderson, J. G. (2000). The Collaborative Longitudinal Personality Disorders Study: Reliability of axis I and II diagnoses. *Journal of Personality Disorders, 14,* 291-299.

Appendix A: Kessler-6 (K6) Module

The Kessler-6 (K6) screening instrument for nonspecific psychological distress (Furukawa, Kessler, Slade, & Andrews, 2003; Kessler et al., 2003) were included in the National Survey on Drug Use and Health (NSDUH) main interview. The two six-item K6 scales gather information regarding how frequently a respondent experienced symptoms of psychological distress during the past 30 days and during a month in the past 12 months, respectively. Only respondents who indicated that there was a worse month (DSTWORST=1) in the past 12 months when he or she felt more depressed, anxious, or emotionally stressed than in the past 30 days were asked about the worst month in the past year scale other than the past 30-day scale.

The questions comprising the two K6 scales and the screener question for the worst month scale are provided below with their associated edited variable names from the mental health module, as well as the response categories for each question:

DSTNRV30 During the past 30 days, how often did you feel nervous?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

Response categories are the same for the remaining past month K6 questions:

DSTHOP30 During the past 30 days, how often did you feel hopeless?

DSTRST30 During the past 30 days, how often did you feel restless or fidgety?

DSTCHR30 During the past 30 days, how often did you feel so sad or depressed that nothing could cheer you up?

DSTEFF30 During the past 30 days, how often did you feel that everything was an effort?

DSTNGD30 During the past 30 days, how often did you feel down on yourself, no good, or worthless?

DSTWORST The last questions asked about how you have been feeling during the past 30 days. Now think about **the past 12 months.** Was there a month in the past 12 months when you felt more depressed, anxious, or emotionally stressed than you felt during the past 30 days?

1 Yes
2 No

DSTNRV12 Think of one month in the past 12 months when you were the most depressed, anxious, or emotionally stressed.

During that month, how often did you feel nervous?

1 All of the time
2 Most of the time
3 Some of the time
4 A little of the time
5 None of the time
DK/REF

Response categories are the same for the remaining worst month K6 questions:

DSTHOP12 During that same month when you were at your worst emotionally . . . how often did you feel hopeless?

DSTRST12 During that same month when you were at your worst emotionally . . . how often did you feel restless or fidgety?

DSTCHR12 During that same month when you were at your worst emotionally . . . how often did you feel so sad or depressed that nothing could cheer you up?

DSTEFF12 During that same month when you were at your worst emotionally . . . how often did you feel that everything was an effort?

DSTNGD12 During that same month when you were at your worst emotionally . . . how often did you feel down on yourself, no good, or worthless?

Each K6 scale item shown above was transformed so that "All of the time" was coded 4, "Most of the time" was coded 3, "Some of the time" was coded 2, "A little of the time" was coded 1, and "None of the time" was coded 0, along with responses matching "Don't know," refusals, bad data, blanks, and legitimate skips.

A past month K6 total score (K6SCMON) was calculated by summing these transformed values across the six past 30-day variables (DSTNRV30, DSTHOP30, DSTRST30, DSTCHR30, DSTEFF30, and DSTNGD30) to arrive at a value ranging between 0 and 24. Likewise, a worst month in the past year K6 total score (K6SCYR) was calculated by summing the transformed values across the six variables for worst month in the past year (DSTNRV12, DSTHOP12, DSTRST12, DSTCHR12, DSTEFF12, and DSTNGD12) to arrive at a value ranging between 0 and 24. The worst month in the past year K6 total score (K6SCYR) has values only for adult respondents who indicated that there was a month in the past year that was worse than the past 30 days (DSTWORST=1). The worst K6 total score in the past year (K6SCMAX) was then created that takes on the maximum value of the past month K6 total score (K6SCMON) and the worst month in the past year K6 total score (K6SCYR) in order to represent the worst K6 total score during the past year, regardless of whether this contradicts the response to DSTWORST.

An alternative past year K6 total score (WSPDSC2) was formulated as follows: past year K6 total scores (K6SCMAX) of less than 8 were recoded as 0, and past year K6 total scores (K6SCMAX) from 8 to 24 were recoded as 1 to 17. The reason for the recoded version was that the serious mental illness (SMI) prevalence was typically extremely low for respondents with

past year K6 total scores of less than 8, and the SMI prevalence rates were higher, in general, only for total scores of 8 or greater. Hence, this variable was used in both the 2008 and 2012 models to create the mental illness variables.

Appendix B: World Health Organization Disability Assessment Schedule (WHODAS) Module

The World Health Organization Disability Assessment Schedule (WHODAS) is a scale used to measure functional impairment that consists of a series of items that are used for assessing disturbances in social adjustment and behavior (i.e., functional impairment). A reduced set of 13 WHODAS items (Novak, Colpe, Barker, & Gfroerer, 2010; Rehm et al., 1999) were administered to a random half of the adult sample in the 2008 National Survey on Drug Use and Health (NSDUH) and to the total adult sample in subsequent NSDUHs (starting in 2009). Responses to this impairment scale were used to create eight variables that were transformed and summed to define the WHODAS total score used in the development of both the 2008 and 2012 serious mental illness (SMI) prediction models. The questions comprising the abbreviated WHODAS are provided below with their associated edited variable names from the mental health module, as well as the response categories for each question:

The next questions are about how much your emotions, nerves, or mental health caused you to have **difficulties in daily activities**.

In answering, think of the **one month** in the past 12 months when your emotions, nerves, or mental health interfered **most** with your daily activities.

IMPREMEM During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **remembering to do things you needed to do**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

IMPCONCN how much difficulty did you have **concentrating on doing something important when other things were going on around you**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

IMPGOUT how much difficulty did you have **going out of the house and getting around on your own**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't leave the house on your own
DK/REF

IMPGOUTM [IF IMPGOUT = 5] Did problems with your emotions, nerves, or mental health keep you from leaving the house on your own?

1 Yes
2 No
DK/REF

IMPPEOP how much difficulty did you have **dealing with people you did not know well**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't deal with people you did not know well
DK/REF

IMPPEOPM [IF IMPPEOP = 5] Did problems with your emotions, nerves, or mental health keep you from dealing with people you did not know well?

1 Yes
2 No
DK/REF

IMPSOC how much difficulty did you have **participating in social activities, like visiting friends or going to parties**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't participate in social activities
DK/REF

IMPSOCM [IF IMPSOC=5] Did problems with your emotions, nerves, or mental health keep you from participating in social activities?

1 Yes
2 No
DK/REF

IMPHHLD how much difficulty did you have **taking care of household responsibilities**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't take care of household responsibilities
DK/REF

IMPHHLDM [IF IMPHHLD=5] Did problems with your emotions, nerves, or mental health keep you from taking care of household responsibilities?

1 Yes
2 No
DK/REF

IMPRESP how much difficulty did you have **taking care of your daily responsibilities at work or school**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
5 You didn't work or go to school
DK/REF

IMPRESPM [IF IMPRESP=5] Did problems with your emotions, nerves, or mental health keep you from working or going to school?

1 Yes
2 No
DK/REF

IMPWORK [IF **IMPRESP** NE 5] During that one month when your emotions, nerves or mental health interfered **most** with your daily activities . . .

how much difficulty did you have **getting your daily work done as quickly as needed**?

1 No difficulty
2 Mild difficulty
3 Moderate difficulty
4 Severe difficulty
DK/REF

An original WHODAS total score (WHODASC2) was created to indicate the level of difficulty in performing daily activities due to problems with emotions, nerves, or mental health. Each of the eight variables created from the WHODAS items shown above was transformed into values of 0 to 3 so that a response of "severe difficulty" was coded 3, "moderate difficulty" was coded 2, "mild difficulty" was coded 1, and "no difficulty" was coded 0, with "don't know" and "refuse" also coded 0. Some items had a fifth category to deal with "not applicable" responses. For example, the question about difficulties regarding taking care of daily responsibilities at work or school (impresp) had a fifth category, "you didn't go to work or school." If this category was selected, then another question was asked as to whether respondents' emotions, nerves, or mental health caused them to be unable to go to work or school (imprespm). A "yes" response to the follow-up question (imprespm=1) was coded 3, and a "no" response (imprespm=2) was coded 0. One exception to this coding was the last WHODAS recode on how much difficulty the respondents had in getting their daily work done as quickly as needed (impwork). This item was asked of the respondents only if in the previous question they responded that they went to work or school (impresp=1 to 4). In the case that they responded that they did not go to work or school (impresp=5), their response to the follow-up question referred to above (imprespm) determined the final item score for impwork; otherwise, impwork was recoded similar to the other items.

The transformed scale values were summed across the eight variables created from the WHODAS items (remembering, concentrating, going out of the house on your own, dealing with people you don't know well, participating in social activities, taking care of household responsibilities, taking care of daily work/school responsibilities, and getting your daily work done as quickly as needed) to arrive the WHODAS total score (WHODASC2) at a value ranging between 0 and 24.

An alternative WHODAS total score (WHODASC3) was created to indicate the number of daily activities in which a respondent had moderate or severe difficulty performing or did not perform due to problems with emotions, nerves, or mental health. Each of the eight variables created from WHODAS items shown above was transformed into values of 0 or 1 so that responses indicating "moderate difficulty" or "severe difficulty" were recoded 1 and responses indicating "mild difficulty" or "no difficulty" were recoded 0. If a fifth category of "not applicable" was available and selected (see above for an example of a fifth question), then another question was asked as to whether respondents' emotions, nerves, or mental health caused them to be unable to go to work or school. A "yes" response was recoded 1, and a "no" response was recoded 0. The transformed scale values were summed across the eight WHODAS activities to arrive at a value ranging between 0 and 8. The recoded version of the WHODAS total score was driven by the idea that a dichotomous measure dividing respondents who experienced moderate or severe difficulties from the remaining respondents might fit better than a linear continuous measure. This alternative WHODAS total score was used in both the 2008 and 2012 models for estimating the mental illness variables.

Appendix C: Sheehan Disability Scale (SDS) Module

The Sheehan Disability Scale (SDS) is a scale used to measure functional impairment that consists of four questions that ask respondents how much their emotions, nerves, or mental health interfered with their daily activities over the past year (Leon, Olfson, Portera, Farber, & Sheehan, 1997). It was administered to a random half of the adult sample in the 2008 National Survey on Drug Use and Health (NSDUH). The questions comprising the SDS are provided below with their associated edited variable names from the mental health module, as well as the response categories for each question:

The next questions are about how much your emotions, nerves, or mental health **interfered with your daily activities**. In answering, think of the **one month** in the past 12 months when your emotions, nerves, or mental health interfered **most** with your daily activities. Using the 0 to 10 scale shown below, where 0 means **no** interference and 10 means very **severe** interference, select the number that describes how much **your emotions, nerves or mental health** interfered with your **ability to do** each of the following activities during that period. You can use any number between 0 and 10 to answer.

MHSMGT **During that month when you were at your worst emotionally,** how much did your emotions interfere with your **ability to do home management tasks**, like cleaning, shopping, and working around the house, apartment, or yard?

Response categories are the same for the remaining SDS questions:

MHSWORK During that **month** in the past 12 months when you were at your worst emotionally how much did this interfere with your **ability to work**?

MHSRELS During that month when you were at your worst emotionally, how much did this interfere with your **ability to form and maintain close relationships** with other people?

MHSSOC How much did your emotions interfere with your **ability to have a social life** during that period of time?

The original SDS total score (SDSSC2) was created to indicate the level of interference with daily activities due to problems with emotions, nerves, or mental health. Each SDS item shown above was transformed so that responses indicating "very severe interference" were coded as 10; responses indicating "severe interference" were coded as 7, 8, and 9; responses indicating "moderate interference" were coded as 4, 5, and 6; responses indicating "mild interference" were coded as 1, 2, and 3; and responses indicating "no interference" were coded as 0. Additionally, sample B respondents who did not answer the questions or those with unknown data were coded as 0. Youths aged 12 to 17 and those adult respondents not in sample B or not assigned to a subsample were assigned the standard SAS missing code (.). The transformed scale values were summed across the four SDS activities (home management, ability to work, ability to form and

maintain close relationships, and ability to have a social life) to arrive at a value ranging between 0 and 40.

The alternative SDS total score (SDSSC3) indicates the number of daily activities in which emotions, nerves, or mental health severely or very severely interfered with performing the activity in the one month in the past year when they were at their worst emotionally. Each SDS item shown above was transformed so that responses indicating "severe interference" or "very severe interference" were recoded as 1. Responses indicating "no interference," "mild interference," or "moderate interference" and sample B respondents who did not answer the questions or those with unknown data were recoded as 0. The transformed scale values were summed across the four SDS activities (see above) to arrive at a value ranging between 0 and 4. The recoded version of the SDS total score also was driven by the idea that a dichotomous measure dividing respondents who experienced severe or very severe interference from the remaining respondents might fit better than a linear continuous measure.

Appendix D: Descriptive Characteristics for Undercoverage of Hispanics, Agreement to Participate, and Completion of the Clinical Follow-Up

Table D1a Descriptive Characteristics, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH

Variable	Hispanics Responding in English Estimate (%)	Hispanics Responding in Spanish Estimate (%)
GENDER		
Male	51.8[a]	49.7
Female	48.2[a]	50.3
AGE		
18 to 25	25.2[c]	10.5
26 to 34	23.7[b]	21.6
35 to 49	28.6[c]	36.0
50 or Older	22.5[c]	31.9
AGE		
18 to 30	39.3[c]	22.1
31 or older	60.7[c]	77.9
COUNTY TYPE		
Large Metro	64.6[c]	73.0
Small Metro	28.0[c]	22.4
Nonmetro	7.4[c]	4.5
EDUCATION		
Less than High School	21.3[c]	62.5
High School Graduate	32.8[c]	22.2
Some College	28.2[c]	9.4
College Graduate	17.7[c]	5.9
MARITAL STATUS		
Married	45.5[c]	57.8
Widowed	2.5[c]	5.2
Divorced or Separated	14.2[b]	12.4
Never Married	37.8[c]	24.6
INCOME		
Less than $20,000	20.0[c]	37.7
$20,000 to $49,999	39.5[c]	49.0
$50,000 to $74,999	16.3[c]	8.1
$75,000 or More	24.3[c]	5.3
REGION		
Northeast	15.7[c]	12.5
Midwest	9.3[c]	7.5
South	36.8	34.5
West	38.2[c]	45.5
NOT CURRENTLY WORKING BUT WORKED IN THE PAST 12 MONTHS[1]		
Yes	9.1[c]	7.4
Other[2]	90.9[c]	92.6
RETIREMENT STATUS IN PAST WEEK		
No Job: Retired	6.7	8.1
Other[3]	93.3	91.9
WORK SITUATION IN PAST WEEK		
No Job: Looking for Work/Layoff, Not Looking for Work/Disabled for Work	12.5	11.3
Other[4]	87.5	88.7
HEALTH INSURANCE[5]		
Covered by Any Health Insurance	71.1[c]	45.2
Not Covered by Any Health Insurance	28.9[c]	54.8
HAD DEPRESSION IN LIFETIME[6]		
Yes	8.9[c]	5.3
No	91.1[c]	94.7

(continued)

Table D1a Descriptive Characteristics, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH (continued)

Variable	Hispanics Responding in English	Hispanics Responding in Spanish
	Estimate (%)	Estimate (%)
HAD ULCER(S) IN LIFETIME[6]		
Yes	2.0	2.3
No	98.0	97.7
YEARS LIVED IN THE U.S.		
Less than 5 years	1.5[c]	12.5
At least 5 years but less than 10 years	3.3[c]	18.8
10 years or more	25.8[c]	64.5
Born in U.S.	69.4[c]	4.1
INTERVIEW PRIVACY		
Completely private	82.1	80.6
Minor distractions	13.1[b]	14.9
Person(s) in the room or listening about 1/3 time	2.1	2.0
Serious interruptions of privacy more than 1/2 time	0.4	0.4
Constant presence of other person(s)	2.4	2.1
INTERVIEW LENGTH: Adult Depression		
Less than 2 minutes[7]	83.5[c]	91.3
Greater or equal to 2 minutes	16.5[c]	8.7
INTERVIEW LENGTH: Entire Interview		
Less than 60 minutes[7]	39.9[c]	13.8
Greater or equal to 60 minutes	60.1[c]	86.2
NO DIFFICULTY UNDERSTANDING THE MAIN NSDUH INTERVIEW		
No difficulty	88.3[c]	78.7
Other[8]	11.7[c]	21.3

NSDUH = National Survey on Drug Use and Health.

* Low precision.

NOTE: For 2008, the 10 cases that were in neither sample (A or B) have been excluded.

[a] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.10$.

[b] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.05$.

[c] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.01$.

[1] This question only includes respondents who said DK/REF to having worked in the last week or NO/DK/REF to DID YOU HAVE JOB BUSINESS in the past week.

[2] Other includes legitimate skip, did not work in the past 12 months, or missing.

[3] Other includes other no job categories, has job, or missing.

[4] Other includes other no job categories, has job, or missing.

[5] Respondents with unknown health insurance status were excluded.

[6] Respondents with unknown health data were excluded.

[7] Missing or negative data are included in this category.

[8] Other includes little difficulty, some difficulty, a lot of difficulty, or missing.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2011.

Table D1b Mental Health Measures, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH

Variable	Hispanics Responding in English Estimate (%)	Hispanics Responding in Spanish Estimate (%)
K6 Total Score		
K6 Total Score: Maximum of Past Year and Past 30 Days (Range: 0-24)[1]	5.1[c]	3.3
Score 0 to 3	51.2[c]	68.3
Score 4 to 5	13.8[c]	9.5
Score 6 to 7	9.7[c]	6.5
Score 8 to 9	6.5[c]	4.7
Score 10 to 11	4.8[c]	3.1
Score 12 to 15	6.7[c]	3.3
Score 16 or higher	7.4[c]	4.6
WHODAS Total Score[2]		
Total Score (Range 0-24)[1]	3.4[c]	1.9
Alternative Total Score (Range 0-8)[1]	0.8[c]	0.4
MDE[3]		
Lifetime MDE	11.6[c]	5.4
Past Year MDE	6.4[c]	3.3
MDE TREATMENT[4]		
Saw or Talk to MD/Professional for MDE in Past Year	53.6	49.5
Used Rx Medication for MDE in Past Year	35.3	*
MENTAL HEALTH TREATMENT		
Received Any Mental Health Treatment in Past Year[5]	9.2[c]	3.9
Stay over in Hospital for Mental Health Treatment in Past 12 Months[5]	1.1[c]	0.6
Received Outpatient Mental Health Treatment in Past 12 Months[5]	5.2[c]	1.7
Needed Mental Health Treatment but Didn't Get It in Past 12 Months	5.0[c]	1.7
SUICIDAL THOUGHTS[6]		
Seriously Think about Killing Self in Past 12 Months	3.6[c]	1.2
Make Plans to Kill Yourself in Past 12 Months	1.0[c]	0.3
Try to Kill Yourself in Past 12 Months	0.6[c]	0.2
SUICIDAL THOUGHTS TREATMENT[6]		
Received Medical Attention Because Tried to Kill Self in Past 12 Months	0.4[c]	0.0
Stay Overnight at Hospital Because Tried to Kill Self in Past 12 Months	0.2[c]	0.0

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; NSDUH = National Survey on Drug Use and Health; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

* Low precision.

NOTE: For 2008, the 10 cases that were in neither sample (A or B) have been excluded. K6SCMAX is used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

[a] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.01$.
[1] Estimates are weighted means.
[2] For 2008, only sample A data are included.
[3] Respondents with unknown lifetime or past year MDE data were excluded.
[4] Among those with past year MDE.
[5] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.
[6] Respondents with unknown suicide information were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2011.

Table D1c Substance Use Measures, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH

Variable	Hispanics Responding in English Estimate (%)	Hispanics Responding in Spanish Estimate (%)
LIFETIME SUBSTANCE USE		
ILLICIT DRUGS[1]	50.1[c]	17.0
Marijuana and Hashish	44.0[c]	8.4
Cocaine	16.8[c]	6.0
Crack	3.5[c]	0.9
Heroin	2.2[c]	0.3
Hallucinogens	15.2[c]	1.1
LSD	8.3[c]	0.2
PCP	2.8[c]	0.1
Ecstasy	7.4[c]	0.4
Inhalants	8.3[c]	1.9
Nonmedical Use of Psychotherapeutics[2,3]	20.7[c]	8.1
Pain Relievers	14.3[c]	6.1
OxyContin®	1.8[c]	0.1
Tranquilizers	7.4[c]	2.1
Stimulants[3]	8.4[c]	1.3
Methamphetamine[3]	5.8[c]	0.8
Sedatives	2.1[c]	0.5
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	32.4[c]	13.7
ALCOHOL	85.8[c]	67.6
TOBACCO[4]	67.7[c]	46.0
CIGARETTES	63.9[c]	44.2
PAST YEAR SUBSTANCE USE		
ILLICIT DRUGS[1]	17.7[c]	5.3
Marijuana and Hashish	13.7[c]	1.4
Cocaine	2.6[c]	1.1
Crack	0.3[b]	0.1
Heroin	0.3[b]	0.1
Hallucinogens	2.1[c]	0.2
LSD	0.3[c]	0.0
PCP	0.1[c]	0.0
Ecstasy	1.6[c]	0.1
Inhalants	0.7[c]	0.2
Nonmedical Use of Psychotherapeutics[2,3]	6.5[c]	3.3
Pain Relievers	5.1[c]	2.5
OxyContin®	0.4[c]	*
Tranquilizers	1.8[c]	0.8
Stimulants[3]	1.1[c]	0.3
Methamphetamine[3]	0.5[c]	0.1
Sedatives	0.3[a]	0.1
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	9.1[c]	4.4
ALCOHOL	72.2[c]	46.4
TOBACCO[4]	34.5[c]	19.9
CIGARETTES	30.5[c]	19.2

(continued)

Table D1c Substance Use Measures, by Language of Interview: Hispanic Respondents Aged 18 or Older, 2008-2011 NSDUH (continued)

Variable	Hispanics Responding in English	Hispanics Responding in Spanish
	Estimate (%)	Estimate (%)
PAST YEAR SUBSTANCE ABUSE OR DEPENDENCE[5]		
ALCOHOL ABUSE OR DEPENDENCE	9.6[c]	5.7
PAIN RELIEVER ABUSE OR DEPENDENCE	0.6	0.4

LSD = lysergic acid diethylamide; NSDUH = National Survey on Drug Use and Health; PCP = phencyclidine.

* Low precision.

NOTE: For 2008, the 10 cases that were in neither sample (A or B) have been excluded.

[a] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.10$.

[b] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.05$.

[c] Difference between this estimate and corresponding estimate from Hispanics Responding in Spanish is statistically significant at $p < 0.01$.

[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. The estimates for Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine incorporated in these summary estimates do not include data from the methamphetamine items added in 2005 and 2006. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.

[2] Nonmedical use of prescription-type psychotherapeutics includes the nonmedical use of pain relievers, tranquilizers, stimulants, or sedatives and does not include over-the-counter drugs.

[3] Estimates of Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine in the designated rows include data from methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.

[4] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.

[5] Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2008-2011.

Table D2a Descriptive Characteristics, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS

Variable	Persons Agreeing to Participate Estimate (%)	Initial Nonrespondents Estimate (%)
GENDER		
Male	48.3	47.9
Female	51.7	52.1
AGE		
18 to 25	18.4[c]	8.6
26 to 34	16.1[c]	8.6
35 to 49	27.2	31.2
50 or Older	38.3[b]	51.6
AGE		
18 to 30	28.3[c]	13.5
31 or older	71.7[c]	86.5
COUNTY TYPE		
Large Metro	54.8	61.9
Small Metro	30.1[b]	22.3
Nonmetro	15.1	15.8
RACE/ETHNICITY		
White	65.9	68.0
Black	10.2[c]	4.6
Hispanic	17.4	*
Asian and Others	6.5	7.7
EDUCATION		
Less than High School	14.2	*
High School Graduate	33.0	39.0
Some College	26.1[c]	17.9
College Graduate	26.8	27.5
MARITAL STATUS		
Married	51.1[b]	63.8
Widowed	4.1	5.8
Divorced or Separated	14.6	13.1
Never Married	30.3[c]	17.3
INCOME		
Less than $20,000	17.2	11.9
$20,000 to $49,999	33.3	37.0
$50,000 to $74,999	18.6	14.6
$75,000 or More	30.9	36.5
REGION		
Northeast	21.5	22.5
Midwest	21.8	18.0
South	31.7	36.4
West	25.0	23.1
NOT CURRENTLY WORKING BUT WORKED IN THE PAST 12 MONTHS[1]		
Yes	8.1[c]	3.8
Other[2]	91.9[c]	96.2
RETIREMENT STATUS IN PAST WEEK		
No Job: Retired	10.7[b]	22.3
Other[3]	89.3[b]	77.7
WORK SITUATION IN PAST WEEK		
No Job: Looking for Work/Layoff, Not Looking for Work/ Disabled for Work	12.6[c]	3.9
Other[4]	87.4[c]	96.1

(continued)

Table D2a Descriptive Characteristics, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS (continued)

Variable	Persons Agreeing to Participate Estimate (%)	Initial Nonrespondents Estimate (%)
HEALTH INSURANCE[5]		
Covered by Any Health Insurance	81.4	85.9
Not Covered by Any Health Insurance	18.6	14.1
HAD DEPRESSION IN LIFETIME[6]		
Yes	15.1[c]	5.9
No	84.9[c]	94.1
HAD ULCER(S) IN LIFETIME[6]		
Yes	5.6[b]	0.9
No	94.4[b]	99.1
YEARS LIVED IN THE US		
Less than 5 years	1.7[a]	0.5
At least 5 years but less than 10 years	2.0	4.1
10 years or more	12.7	*
Born in US	83.6	77.6
INTERVIEW PRIVACY		
Completely private	87.8	86.7
Minor distractions	8.9	*
Person(s) in the room or listening about 1/3 time	1.3	0.8
Serious interruptions of privacy more than 1/2 time	0.2	0.3
Constant presence of other person(s)	1.8	1.7
INTERVIEW LENGTH: Adult Depression		
Less than 2 minutes[7]	77.2[c]	92.5
Greater or equal to 2 minutes	22.8[c]	7.5
INTERVIEW LENGTH: Entire Interview		
Less than 60 minutes[7]	27.7[c]	49.9
Greater or equal to 60 minutes	72.3[c]	50.1
NO DIFFICULTY UNDERSTANDING THE MAIN NSDUH INTERVIEW		
No difficulty	90.3	88.0
Other[8]	9.7	12.0

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

* Low precision.

NOTE: Table subset to only those cases where MHSELECT=1.

[a] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.01$.
[1] This question only includes respondents who said DK/REF to having worked in the last week or NO/DK/REF to DID YOU HAVE JOB BUSINESS in the past week.
[2] Other includes legitimate skip, did not work in the past 12 months, or missing.
[3] Other includes other no job categories, has job, or missing.
[4] Other includes other no job categories, has job, or missing.
[5] Respondents with unknown health insurance status were excluded.
[6] Respondents with unknown health data were excluded.
[7] Missing or negative data are included in this category.
[8] Other includes little difficulty, some difficulty, a lot of difficulty, or missing.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table D2b Mental Health Measures, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS

Variable	Persons Agreeing to Participate Estimate (%)	Initial Nonrespondents Estimate (%)
K6 Total Score		
K6 Total Score: Maximum of Past Year and Past 30 Days (Range: 0-24)[1]	5.2[c]	3.2
Score 0 to 3	51.6[c]	66.9
Score 4 to 5	12.8	12.8
Score 6 to 7	10.6[b]	6.5
Score 8 to 9	6.3	4.5
Score 10 to 11	4.1[b]	2.6
Score 12 to 15	7.3[c]	3.6
Score 16 or higher	7.2[c]	3.1
WHODAS Total Score[2]		
Total Score (Range 0-24)[1]	3.6[c]	2.4
Alternative Total Score (Range 0-8)[1]	0.9[c]	0.6
MDE[3]		
Lifetime MDE	14.7[c]	5.4
Past Year MDE	7.4[c]	3.1
MDE TREATMENT[4]		
Saw or Talk to MD/Professional for MDE in Past Year	64.9	*
Used Rx Medication for MDE in Past Year	58.2	*
MENTAL HEALTH TREATMENT		
Received Any Mental Health Treatment in Past Year[5]	15.2[c]	8.7
Stay over in Hospital for Mental Health Treatment in Past 12 Months[5]	0.7[c]	0.1
Received Outpatient Mental Health Treatment in Past 12 Months[5]	7.7[a]	4.9
Needed Mental Health Treatment but Didn't Get It in Past 12 Months	5.7[c]	1.8
SUICIDAL THOUGHTS[6]		
Seriously Think about Killing Self in Past 12 Months	4.2[c]	1.3
Make Plans to Kill Yourself in Past 12 Months	1.1[c]	0.3
Try to Kill Yourself in Past 12 Months	0.4	0.2
SUICIDAL THOUGHTS TREATMENT[6]		
Received Medical Attention Because Tried to Kill Self in Past 12 Months	0.3[a]	0.1
Stay Overnight at Hospital Because Tried to Kill Self in Past 12 Months	0.2	0.1

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHHS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = 8-item World Health Organization Disability Assessment Schedule.
*Low precision.

NOTE: Table subset to only those cases where MHSELECT=1. K6SCMAX is used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

[a] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.01$.
[1] Estimates are weighted means.
[2] For 2008, only sample A data are included.
[3] Respondents with unknown lifetime or past year MDE data were excluded.
[4] Among those with past year MDE.
[5] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.
[6] Respondents with unknown suicide information were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table D2c Substance Use Measures, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS

Variable	Persons Agreeing to Participate Estimate (%)	Initial Nonrespondents Estimate (%)
LIFETIME		
ILLICIT DRUGS[1]	57.1[c]	38.7
Marijuana and Hashish	53.0[c]	35.4
Cocaine	19.6[c]	10.4
Crack	4.1[c]	1.5
Heroin	2.7[c]	0.8
Hallucinogens	20.8[c]	6.6
LSD	14.3[c]	3.8
PCP	3.1[c]	0.8
Ecstasy	7.7[c]	1.6
Inhalants	11.1[c]	4.3
Nonmedical Use of Psychotherapeutics[2,3]	24.9[c]	13.7
Pain Relievers	18.9[c]	6.1
OxyContin®	3.2[b]	1.5
Tranquilizers	12.9	7.8
Stimulants[3]	11.3[c]	3.3
Methamphetamine[3]	7.7[c]	1.6
Sedatives	4.3[c]	1.2
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	35.5[c]	23.0
ALCOHOL	88.5[a]	78.8
TOBACCO[4]	74.3[b]	61.8
CIGARETTES	67.6	59.5
PAST YEAR		
ILLICIT DRUGS[1]	19.2[c]	6.8
Marijuana and Hashish	15.1[c]	4.3
Cocaine	2.7[c]	0.4
Crack	0.5[c]	0.0
Heroin	0.3[c]	0.0
Hallucinogens	1.8[c]	0.6
LSD	0.3[b]	0.0
PCP	0.0	*
Ecstasy	1.3[c]	0.3
Inhalants	0.4[a]	0.1
Nonmedical Use of Psychotherapeutics[2,3]	9.4[c]	3.5
Pain Relievers	5.7[c]	2.5
OxyContin®	0.8	0.6
Tranquilizers	4.6	1.8
Stimulants[3]	3.3	0.3
Methamphetamine[3]	*	0.1
Sedatives	0.5[a]	0.1
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	11.8[c]	4.2
ALCOHOL	71.7	69.2
TOBACCO[4]	36.7[b]	26.5
CIGARETTES	31.1[b]	22.8

(continued)

Table D2c Substance Use Measures, by Agreement to Participate in the Clinical Follow-Up, 2008-2011 MHSS (continued)

Variable	Persons Agreeing to Participate Estimate (%)	Initial Nonrespondents Estimate (%)
PAST YEAR SUBSTANCE ABUSE OR DEPENDENCE[5]		
ALCOHOL ABUSE OR DEPENDENCE	8.3[b]	4.8
PAIN RELIEVER ABUSE OR DEPENDENCE	1.1	0.6

LSD = lysergic acid diethylamide; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; PCP = phencyclidine.

* Low precision.

NOTE: Table subset to only those cases where MHSELECT=1.

[a] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from initial nonrespondents is statistically significant at $p < 0.01$.
[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. The estimates for Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine incorporated in these summary estimates do not include data from the methamphetamine items added in 2005 and 2006. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.
[2] Nonmedical use of prescription-type psychotherapeutics includes the nonmedical use of pain relievers, tranquilizers, stimulants, or sedatives and does not include over-the-counter drugs.
[3] Estimates of Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine in the designated rows include data from methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.
[4] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.
[5] Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table D3a Descriptive Characteristics, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS

Variable	Respondents to Clinical Follow-Up Estimate (%)	Final Nonrespondents Estimate (%)
GENDER		
Male	46.4[c]	60.6
Female	53.6[c]	39.4
AGE		
18 to 25	14.3[b]	22.6
26 to 34	15.3	19.0
35 to 49	25.6	32.8
50 or Older	44.8[c]	*
AGE		
18 to 30	23.0[c]	36.6
31 or older	77.0[c]	63.4
COUNTY TYPE		
Large Metro	52.1	52.2
Small Metro	29.6	35.7
Nonmetro	18.3[b]	12.0
RACE/ETHNICITY		
White	73.3[c]	55.2
Black	8.4	10.5
Hispanic	12.4[b]	*
Asian and Others	5.8	6.5
EDUCATION		
Less than High School	9.4[c]	*
High School Graduate	35.5	30.9
Some College	25.9	21.8
College Graduate	29.2[c]	16.1
MARITAL STATUS		
Married	56.6[a]	44.9
Widowed	4.3	*
Divorced or Separated	14.7	11.7
Never Married	24.4[c]	39.8
INCOME		
Less than $20,000	13.7[b]	22.9
$20,000 to $49,999	33.3	34.8
$50,000 to $74,999	20.5[c]	12.5
$75,000 or More	32.5	29.8
REGION		
Northeast	24.2[a]	17.3
Midwest	24.1[a]	17.1
South	31.0	35.3
West	20.7[a]	30.4
NOT CURRENTLY WORKING BUT WORKED IN THE PAST 12 MONTHS[1]		
Yes	7.7	5.4
Other[2]	92.3	94.6
RETIREMENT STATUS IN PAST WEEK		
No Job: Retired	14.9[a]	7.9
Other[3]	85.1[a]	92.1
WORK SITUATION IN PAST WEEK		
No Job: Looking for Work/Layoff, Not Looking for Work/Disabled for Work	11.4	12.0
Other[4]	88.6	88.0

(continued)

Table D3a Descriptive Characteristics, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS (continued)

Variable	Respondents to Clinical Follow-Up Estimate (%)	Final Nonrespondents Estimate (%)
HEALTH INSURANCE[5]		
Covered by Any Health Insurance	86.0[c]	69.6
Not Covered by Any Health Insurance	14.0[c]	30.4
HAD DEPRESSION IN LIFETIME[6]		
Yes	13.0	13.0
No	87.0	87.0
HAD ULCER(S) IN LIFETIME[6]		
Yes	3.6	*
No	96.4	*
YEARS LIVED IN THE US		
Less than 5 years	1.0	*
At least 5 years but less than 10 years	1.5	3.3
10 years or more	10.5	*
Born in US	87.0	*
INTERVIEW PRIVACY		
Completely private	88.8	85.3
Minor distractions	8.1	10.7
Person(s) in the room or listening about 1/3 time	1.4	0.9
Serious interruptions of privacy more than 1/2 time	0.1	0.2
Constant presence of other person(s)	1.6	2.9
INTERVIEW LENGTH: Adult Depression		
Less than 2 minutes[7]	81.7	77.3
Greater or equal to 2 minutes	18.3	22.7
INTERVIEW LENGTH: Entire Interview		
Less than 60 minutes[7]	31.7	36.3
Greater or equal to 60 minutes	68.3	63.7
NO DIFFICULTY UNDERSTANDING THE MAIN NSDUH INTERVIEW		
No difficulty	93.7[a]	*
Other[8]	6.3[a]	*

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

* Low precision.

NOTE: Table subset to only those cases who initially agreed to participate in the clinical follow-up interview.

[a] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.01$.
[1] This question only includes respondents who said DK/REF to having worked in the last week or NO/DK/REF to DID YOU HAVE JOB BUSINESS in the past week.
[2] Other includes legitimate skip, did not work in the past 12 months, or missing.
[3] Other includes other no job categories, has job, or missing.
[4] Other includes other no job categories, has job, or missing.
[5] Respondents with unknown health insurance status were excluded.
[6] Respondents with unknown health data were excluded.
[7] Missing or negative data are included in this category.
[8] Other includes little difficulty, some difficulty, a lot of difficulty, or missing.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table D3b Mental Health Measures, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS

Variable	Respondents to Clinical Follow-Up Estimate (%)	Final Nonrespondents Estimate (%)
K6 Total Score		
K6 Total Score: Maximum of Past Year and Past 30 Days (Range: 0-24)[1]	4.7	4.9
Score 0 to 3	54.3	56.6
Score 4 to 5	13.6	10.0
Score 6 to 7	11.3[b]	7.4
Score 8 to 9	5.1	8.1
Score 10 to 11	3.8	3.2
Score 12 to 15	6.1	7.1
Score 16 or Higher	5.7	7.6
WHODAS Total Score[2]		
Total Score (Range 0-24)[1]	3.3	3.4
Alternative Total Score (Range 0-8)[1]	0.8	0.9
MDE[3]		
Lifetime MDE	12.6	12.7
Past Year MDE	6.3	6.8
MDE TREATMENT[4]		
Saw or Talk to MD/Professional for MDE in Past Year	67.4[a]	*
Used Rx Medication for MDE in Past Year	59.5	53.1
MENTAL HEALTH TREATMENT		
Received Any Mental Health Treatment in Past Year[5]	13.6	14.0
Stay over in Hospital for Mental Health Treatment in Past 12 Months[5]	0.4[b]	1.2
Received Outpatient Mental Health Treatment in Past 12 Months[5]	6.9	6.4
Needed Mental Health Treatment but Didn't Get It in Past 12 Months	4.5	6.2
SUICIDAL THOUGHTS[6]		
Seriously Think about Killing Self in Past 12 Months	3.5	3.9
Make Plans to Kill Yourself in Past 12 Months	0.9	1.2
Try to Kill Yourself in Past 12 Months	0.2[b]	1.0
SUICIDAL THOUGHTS TREATMENT[6]		
Received Medical Attention Because Tried to Kill Self in Past 12 Months	0.1[b]	0.7
Stay Overnight at Hospital Because Tried to Kill Self in Past 12 Months	0.0[b]	0.6

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHHS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

* Low precision.

NOTE: Table subset to only those cases who initially agreed to participate in the clinical follow-up interview. K6SCMAX is used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

[a] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.01$.
[1] Estimates are weighted means.
[2] For 2008, only sample A data are included.
[3] Respondents with unknown lifetime or past year MDE data were excluded.
[4] Among those with past year MDE.
[5] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.
[6] Respondents with unknown suicide information were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table D3c Substance Use Measures, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS

Variable	Respondents to Clinical Follow-Up Estimate (%)	Final Nonrespondents Estimate (%)
LIFETIME		
ILLICIT DRUGS[1]	53.0	60.4
Marijuana and Hashish	49.5	54.7
Cocaine	16.8	20.0
Crack	2.9[b]	6.0
Heroin	1.6[a]	4.6
Hallucinogens	17.2	21.1
LSD	11.2	14.4
PCP	2.5	2.7
Ecstasy	5.5[a]	9.0
Inhalants	10.1	9.7
Nonmedical Use of Psychotherapeutics[2,3]	20.7	27.9
Pain Relievers	14.6	20.6
OxyContin®	2.0[b]	4.8
Tranquilizers	10.0	12.5
Stimulants[3]	9.6	9.3
Methamphetamine[3]	6.2	6.8
Sedatives	4.3[a]	2.7
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	31.0	38.1
ALCOHOL	89.6	86.2
TOBACCO[4]	72.5	78.8
CIGARETTES	67.2	*
PAST YEAR		
ILLICIT DRUGS[1]	15.5	19.1
Marijuana and Hashish	12.2	14.4
Cocaine	2.1	2.6
Crack	0.2[c]	0.8
Heroin	0.1[b]	0.6
Hallucinogens	1.5	1.5
LSD	0.2	0.4
PCP	0.0	0.1
Ecstasy	1.1	1.1
Inhalants	0.3	0.4
Nonmedical Use of Psychotherapeutics[2,3]	6.9	9.9
Pain Relievers	4.2	7.5
OxyContin®	0.4[b]	1.3
Tranquilizers	3.1	4.0
Stimulants[3]	2.3	1.9
Methamphetamine[3]	*	0.9
Sedatives	0.3	0.5
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	9.1	11.2
ALCOHOL	73.2[a]	61.8
TOBACCO[4]	33.2[b]	43.2
CIGARETTES	26.3[b]	39.3

(continued)

Table D3c Substance Use Measures, by Participation in Clinical Interview: Persons Initially Agreeing to Participate in Clinical Follow-Up, 2008-2011 MHSS (continued)

Variable	Respondents to Clinical Follow-Up Estimate (%)	Final Nonrespondents Estimate (%)
PAST YEAR SUBSTANCE ABUSE OR DEPENDENCE[5]		
ALCOHOL ABUSE OR DEPENDENCE	6.2[b]	12.4
PAIN RELIEVER ABUSE OR DEPENDENCE	0.5[b]	2.1

LSD = lysergic acid diethylamide; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; PCP = phencyclidine.

* Low precision.

NOTE: Table subset to only those cases who initially agreed to participate in the clinical follow-up interview.

[a] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.10$.
[b] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.05$.
[c] Difference between this estimate and corresponding estimate from final nonrespondents is statistically significant at $p < 0.01$.
[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. The estimates for Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine incorporated in these summary estimates do not include data from the methamphetamine items added in 2005 and 2006. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.
[2] Nonmedical use of prescription-type psychotherapeutics includes the nonmedical use of pain relievers, tranquilizers, stimulants, or sedatives and does not include over-the-counter drugs.
[3] Estimates of Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine in the designated rows include data from methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.
[4] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.
[5] Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Appendix E: Descriptive Characteristics for Overall Impact of Weighting Adjustments

Table E1 Descriptive Characteristics, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample

		Selected MHSS Clinical Sample				
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)
GENDER						
Male	48.3	48.2	52.8[b2]	45.1[a3]	48.3	48.3
Female	51.7	51.8	47.2[b2]	54.9[a3]	51.7	51.7
AGE						
18 to 25	14.8	16.2	15.6	16.6[a3]	14.8	14.8
26 to 34	15.8	14.4[a1]	12.7[b1a2]	15.5	17.2	15.5
35 to 49	27.5	28.1	32.9[b2]	24.9	26.1[a4]	27.8
50 or Older	41.9	41.3	38.8	43.0	41.9	41.9
AGE						
18 to 30	23.9	25.0	24.0	25.6	24.9	23.3
31 or Older	76.1	75.0	76.0	74.4	75.1	76.7
COUNTY TYPE						
Large Metro	53.2	56.4[a1]	59.1[a1]	54.6[a3]	50.8	52.0
Small Metro	30.6	28.3	26.2[b1]	29.7	31.8	30.7
Nonmetro	16.1	15.3	14.6	15.7	17.4	17.4
RACE/ETHNICITY						
White	68.0	66.4	59.9[b1b2]	70.8	68.0	68.0
Black	11.5	8.9[c1]	7.9[c1]	9.6[c3]	11.5	11.4
Hispanic	13.9	17.9[a1]	24.9[c1b2]	13.2	13.9	13.9
Asian and Others	6.6	6.8	7.3	6.4	6.6	6.7
EDUCATION						
Less than High School	15.1	14.5	22.3[a1c2]	9.3[c1b3]	12.1[a1c4]	8.2[c1]
High School Graduate	30.5	34.3[a1]	35.3	33.6[a3]	30.0	32.5
Some College	25.8	24.2	19.7[c1c2]	27.3	27.3	28.0
College Graduate	28.6	26.9	22.7[b1b2]	29.8	30.6	31.3
MARITAL STATUS						
Married	53.8	54.0	53.9	54.0	56.0	55.1
Widowed	6.0	4.5[b1]	5.2	4.0[b1]	3.4[c1]	3.4[c1]
Divorced or Separated	13.7	14.2	12.8	15.2	15.1	15.3
Never Married	26.5	27.3	28.1	26.8	25.6	26.2
INCOME						
Less than $20,000	18.2	16.0	18.2	14.5	13.8[c1a4]	12.2[c1]
$20,000 to $49,999	32.8	34.1	36.9	32.3	33.0	34.0
$50,000 to $74,999	17.4	17.7	13.4[c1b2]	20.6	19.6	19.6
$75,000 or More	31.6	32.2	31.4	32.7	33.5	34.2
REGION						
Northeast	18.5	21.8[b1]	20.4	22.7[a1]	22.5[a1]	21.3[a1]
Midwest	21.8	21.0	17.3[b1b2]	23.4	23.2	23.1
South	36.6	32.7[b1]	34.9	31.2[b1]	33.0[a1]	33.2
West	23.1	24.5	27.4	22.6	21.3	22.5
NOT CURRENTLY WORKING BUT WORKED IN THE PAST 12 MONTHS[1]						
Yes	7.8	7.1	4.9[c1b2]	8.6[a3]	7.3	7.7
Other[2]	92.2	92.9	95.1[c1b2]	91.4[a3]	92.7	92.3
RETIREMENT STATUS IN PAST WEEK						
No Job: Retired	14.9	13.3	15.8	11.6[b1]	12.6	12.0[a1]
Other[3]	85.1	86.7	84.2	88.4[b1]	87.4	88.0[a1]
WORK SITUATION IN PAST WEEK						
No Job: Looking for Work/ Layoff, Not Looking for Work/Disabled for Work	10.4	10.6	8.1[b1b2]	12.3	11.2	10.5
Other[4]	89.6	89.4	91.9[b1b2]	87.7	88.8	89.5

(continued)

Table E1 Descriptive Characteristics, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample (continued)

		Selected MHSS Clinical Sample				
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)	Estimate (%)
HEALTH INSURANCE[5]						
Covered by Any Health Insurance	82.1	82.4	77.6[b2]	85.7[c1]	85.3[c1]	84.9[a1]
Not Covered by Any Health Insurance	17.9	17.6	22.4[b2]	14.3[c1]	14.7[c1]	15.1[a1]
HAD DEPRESSION IN LIFETIME[6]						
Yes	12.6	13.0	10.3[a1b2]	14.8[b1a3]	13.4[c4]	15.3[c1]
No	87.4	87.0	89.7[a1b2]	85.2[b1a3]	86.6[c4]	84.7[c1]
HAD ULCER(S) IN LIFETIME[6]						
Yes	3.3	4.5	*	4.2	3.7[c4]	4.5
No	96.7	95.5	*	95.8	96.3[c4]	95.5
YEARS LIVED IN THE US						
Less than 5 years	1.7	1.4	2.0	1.0[b1]	1.0[b1]	0.9[c1]
At least 5 years but less than 10 years	2.2	2.5	3.9	1.5	1.3[a1]	1.1[c1]
10 years or more	11.7	13.8	17.8	11.2	10.3	9.1[a1]
Born in US	84.4	82.3	76.4[a1b2]	86.3	87.4[a1]	88.8[c1]
INTERVIEW PRIVACY						
Completely private	85.4	87.5[a1]	86.1	88.5[c1]	88.2[b1]	88.0[b1]
Minor distractions	10.6	9.3	10.7	8.3[c1]	8.3[c1]	8.6[c1]
Person(s) in the room or listening about 1/3 time	1.5	1.2	0.9[a1]	1.4	1.6	1.5
Serious interruptions of privacy more than 1/2 time	0.3	0.2	0.3	0.2	0.2	0.2
Constant presence of other person(s)	2.2	1.8	2.1	1.6	1.7	1.8
INTERVIEW LENGTH: Adult Depression						
Less than 2 minutes[7]	81.4	80.7	83.6	78.7[a1b3]	80.8[c4]	77.9[b1]
Greater or equal to 2 minutes	18.6	19.3	16.4	21.3[a1b3]	19.2[c4]	22.1[b1]
INTERVIEW LENGTH: Entire Interview						
Less than 60 minutes[7]	44.0	32.7[c1]	40.2[c2]	27.7[c1b3]	32.4[c1c4]	25.6[c1]
Greater or equal to 60 minutes	56.0	67.3[c1]	59.8[c2]	72.3[c1b3]	67.6[c1c4]	74.4[c1]

(continued)

Table E1 Descriptive Characteristics, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample (continued)

		Selected MHSS Clinical Sample				
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**
NO DIFFICULTY UNDERSTANDING THE MAIN NSDUH INTERVIEW						
No difficulty	90.5	89.8	84.0[a1b2]	93.7[b1]	92.2	93.6[b1]
Other[8]	9.5	10.2	16.0[a1b2]	6.3[b1]	7.8	6.4[b1]

MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health.

* Low precision.

NOTE: For 2008 estimates (column 1), the 10 cases that were in neither sample (A or B) have been excluded; Column 1 weight= ANALWT_A, where ANALWT_A=ANALWT for 2009-2012, ANALWT_A=MHSAMPWT for 2008; Columns 2-4 weight= MHDSNWT, design-based weight; column 5 weight= MHNEWWGT, MHSS final revised weight; Column 6 weight= MHWEIGHT for 2008-2011, and MHSSWT1 for 2008 Sample A, previously adjusted weight.

[a1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.10$.
[b1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.05$.
[c1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.01$.
[a2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.10$.
[b2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.05$.
[c2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.01$.
[a3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.10$.
[b3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.05$.
[c3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.01$.
[a4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.10$.
[b4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.05$.
[c4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.01$.
[1] This question only includes respondents who said DK/REF to having worked in the last week or NO/DK/REF to DID YOU HAVE JOB BUSINESS in the past week.
[2] Other includes legitimate skip, did not work in the past 12 months, or missing.
[3] Other includes other no job categories, has job, or missing.
[4] Other includes other no job categories, has job, or missing.
[5] Respondents with unknown health insurance status were excluded.
[6] Respondents with unknown health data were excluded.
[7] Missing or negative data are included in this category.
[8] Other includes little difficulty, some difficulty, a lot of difficulty, or missing.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table E2 Mental Health, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample

			Selected MHSS Clinical Sample			
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**
K6 Total Score						
K6 Total Score: Maximum of Past Year and Past 30 Days (Range: 0-24)[1]	4.9	4.8	4.2[b1c2]	5.2	5.1	5.3[b1]
Score 0 to 3	53.4	55.0	60.8[b1b2]	51.2	51.1	51.5
Score 4 to 5	13.7	12.8	11.3	13.8	13.4	13.0
Score 6 to 7	9.3	9.7	7.0[b1c2]	11.5[a1]	12.1[b1c4]	10.1
Score 8 to 9	6.1	5.9	6.6	5.4	5.6	6.0
Score 10 to 11	4.5	3.8[b1]	3.0[c1b2]	4.3	4.0	4.9
Score 12 to 15	6.7	6.5	5.7[a2]	7.0	7.6	7.1
Score 16 or higher	6.2	6.3	5.6	6.8[b3]	6.1[c4]	7.4
WHODAS Total Score[2]						
Total Score (Range 0-24)[1]	3.5	3.3	3.0[a1a2]	3.6	3.4[b4]	3.6
Alternative Total Score (Range 0-8)[1]	0.9	0.8	0.8	0.9	0.9	0.9
MDE[3]						
Lifetime MDE	12.8	12.6	9.5[c1c2]	14.7[b3]	12.9[c4]	14.6[a1]
Past Year MDE	6.6	6.4	5.3[b1b2]	7.2	6.6[a4]	7.2
MDE TREATMENT[4]						
Saw or Talk to MD/Professional for MDE in Past Year	61.8	64.5	56.3	68.5[b1a3]	65.9	66.6
Used Rx Medication for MDE in Past Year	51.7	57.7[b1]	53.4	59.7[b1]	56.4	56.9
MENTAL HEALTH TREATMENT						
Received Any Mental Health Treatment in Past Year[5]	13.6	13.7	12.2	14.7	13.7	14.2
Stay over in Hospital for Mental Health Treatment in Past 12 Months[5]	0.8	0.5[a1]	0.6	0.5	0.5	0.4[a1]
Received Outpatient Mental Health Treatment in Past 12 Months[5]	6.6	7.1	6.2	7.7[a1b3]	6.9[a4]	7.9[a1]
Needed Mental Health Treatment but Didn't Get It in Past 12 Months	4.9	4.8	4.2	5.2	4.8	5.4
SUICIDAL THOUGHTS[6]						
Seriously Think about Killing Self in Past 12 Months	3.7	3.5	2.8[c1b2]	4.1	3.7	3.8
Make Plans to Kill Yourself in Past 12 Months	1.0	1.0	0.8	1.0	1.1	1.0
Try to Kill Yourself in Past 12 Months	0.5	0.4	0.7[b2]	0.2[c1]	0.2[c1]	0.2[c1]

(continued)

Table E2 Mental Health, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample (continued)

			Selected MHSS Clinical Sample			
	Full NSDUH Sample (1)	MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**
SUICIDAL THOUGHTS TREATMENT6						
Received Medical Attention Because Tried to Kill Self in Past 12 Months	0.3	0.2	0.4[b2]	0.1[c1]	0.1[c1]	0.1[c1]
Stay Overnight at Hospital Because Tried to Kill Self in Past 12 Months	0.2	0.2	0.4[b2]	0.0[c1]	0.0[c1]	0.0[c1]

K6 = Kessler-6, a 6-item psychological distress scale; MDE = major depressive episode; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; WHODAS = 8-item World Health Organization Disability Assessment Schedule.

* Low precision.

NOTE: For 2008 estimates (column 1), the 10 cases that were in neither sample (A or B) have been excluded; Column 1 weight= ANALWT_A, where ANALWT_A=ANALWT for 2009-2012, ANALWT_A=MHSAMPWT for 2008; Columns 2-4 weight= MHDSNWT, design-based weight; column 5 weight= MHNEWWGT, MHSS final revised weight; Column 6 weight= MHWEIGHT for 2008-2011, and MHSSWT1 for 2008 Sample A, previously adjusted weight. K6SCMAX is used in this table, which is defined as the higher K6 score between the past month K6 total score and the K6 total score in the worst month of the past year if the worst month was not the past 30 days.

[a1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.10$.
[b1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.05$.
[c1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.01$.
[a2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.10$.
[b2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.05$.
[c2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.01$.
[a3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.10$.
[b3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.05$.
[c3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.01$.
[a4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.10$.
[b4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.05$.
[c4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.01$.
[1] Estimates are weighted means.
[2] For 2008, only sample A data are included.
[3] Respondents with unknown lifetime or past year MDE data were excluded.
[4] Among those with past year MDE.
[5] Mental health treatment/counseling is defined as having received inpatient care or outpatient care or having used prescription medication for problems with emotions, nerves, or mental health. Respondents were not to include treatment for drug or alcohol use. Respondents with unknown treatment/counseling information were excluded.
[6] Respondents with unknown suicide information were excluded.

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.

Table E3 Substance Use Measures, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample

			Selected MHSS Clinical Sample			
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**
LIFETIME						
ILLICIT DRUGS[1]	49.5	52.9[a1]	50.3	54.7[b1]	54.7[b1]	55.6[c1]
Marijuana and Hashish	44.4	49.0[b1]	45.8	51.2[c1]	50.5[c1]	52.1[c1]
Cocaine	16.0	17.5	15.1	19.2	17.5	18.2
Crack	3.7	3.5	3.7	3.4	3.4	3.6
Heroin	1.7	2.3	2.8	1.9	1.9	2.2
Hallucinogens	15.7	17.6	14.0[b2]	20.0[a1]	18.0	18.8[b1]
LSD	10.1	11.9	9.7	13.4	11.3	12.5
PCP	2.7	2.6	1.9[a1]	3.0	2.6	3.0
Ecstasy	6.1	6.3	5.7	6.7[a3]	5.6[a4]	6.2
Inhalants	8.6	9.5	7.6[a2]	10.8[a1]	10.3	10.3
Nonmedical Use of Psychotherapeutics[2,3]	21.5	22.4	20.7	23.5	21.5	21.9
Pain Relievers	14.3	16.0	13.5	17.7	14.9	16.2
OxyContin®	2.3	2.8	3.4	2.4	2.1	2.0
Tranquilizers	9.2	11.8[a1]	10.5	12.6[b3]	8.8[c4]	11.2
Stimulants[3]	9.1	9.5	6.3[c1b2]	11.7	9.7	11.0
Methamphetamine[3]	5.5	6.3	4.1[b1]	7.8	6.0	7.2
Sedatives	3.5	3.6	2.1[c1c2]	4.7[a1]	3.8	4.3
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	31.4	32.6	30.7	34.0	32.6	33.0
ALCOHOL	87.4	86.3	81.8[a1a2]	89.4	89.8[a4]	91.8[c1]
TOBACCO[4]	73.4	71.5	69.0	73.1	73.9	75.2
CIGARETTES	68.8	65.8	62.1[a1]	68.3	68.4	70.3
PAST YEAR						
ILLICIT DRUGS[1]	14.4	16.4	13.3[a2]	18.4	16.3	16.9
Marijuana and Hashish	10.9	12.6	9.4[a2]	14.8	12.3	13.5
Cocaine	1.9	2.1	1.6[a2]	2.5	2.0	2.2
Crack	0.4	0.4	0.4	0.3	0.3	0.3
Heroin	0.2	0.3	0.4[a2]	0.2	0.2[a4]	0.2
Hallucinogens	1.5	1.5	1.2[a1]	1.8	1.8	1.6
LSD	0.3	0.3	0.2	0.3	0.3	0.4
PCP	0.0	0.0	0.0	0.0	*	*
Ecstasy	0.9	1.1	0.8	1.3	1.2	1.1
Inhalants	0.5	0.3[a1]	0.3[b1]	0.3	0.4	0.3
Nonmedical Use of Psychotherapeutics[2,3]	6.0	8.1	6.9	8.8	6.7	7.4
Pain Relievers	4.5	5.0	5.3	4.8	5.0	4.4
OxyContin®	0.6	0.7	1.0[a2]	0.5	0.5	0.4[b1]
Tranquilizers	2.1	3.9	3.0	*	2.2	3.5
Stimulants[3]	1.1	2.6	1.0	*	1.5	*
Methamphetamine[3]	0.4	*	0.5	*	0.9	*
Sedatives	0.3	0.4	0.3	0.4[a3]	0.3[b4]	0.4
ILLICIT DRUGS OTHER THAN MARIJUANA[1]	7.6	10.1	8.1	11.5	9.0	10.0
ALCOHOL	70.3	71.2	67.1	73.9	73.2	75.1[c1]
TOBACCO[4]	34.5	34.4	34.3	34.4	34.2	34.6
CIGARETTES	28.6	29.2	30.6	28.3	28.5	28.1

(continued)

Table E3 Substance Use Measures, by Sample and Response Status among Persons Aged 18 or Older, 2008-2011 Adult NSDUH Main Study and Clinical Sample (continued)

			Selected MHSS Clinical Sample			
	Full NSDUH Sample (1)	Selected MHSS Clinical Sample (2)	MHSS Nonrespondents (3)	Respondents before Adjustment (4)	Respondents after Adjustment (Revised Weights) (5)	Respondents after Adjustment (Previous Weights) (6)
Variable	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**	**Estimate (%)**
PAST YEAR SUBSTANCE ABUSE OR DEPENDENCE[5]						
ALCOHOL ABUSE OR DEPENDENCE	7.4	7.5	8.3	7.0	7.3	6.9
PAIN RELIEVER ABUSE OR DEPENDENCE	0.7	1.0	1.5[a1b2]	0.7	0.7[a4]	0.6

LSD = lysergic acid diethylamide; MHSS = Mental Health Surveillance Study; NSDUH = National Survey on Drug Use and Health; PCP = phencyclidine.

* Low precision.

NOTE: For 2008 estimates (column 1), the 10 cases that were in neither sample (A or B) have been excluded; Column 1 weight= ANALWT_A, where ANALWT_A=ANALWT for 2009-2012, ANALWT_A=MHSAMPWT for 2008; Columns 2-4 weight= MHDSNWT, design-based weight; column 5 weight= MHNEWWGT, MHSS final revised weight; Column 6 weight= MHWEIGHT for 2008-2011, and MHSSWT1 for 2008 Sample A, previously adjusted weight.

[a1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.10$.

[b1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.05$.

[c1] Difference between this estimate (from columns 2-6) and corresponding estimate from full CAI sample (column 1) is statistically significant at $p < 0.01$.

[a2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.10$.

[b2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.05$.

[c2] Difference between this estimate from nonrespondents (column 3) and corresponding estimate from respondents (column 4) using design-based weights is statistically significant at $p < 0.01$.

[a3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.10$.

[b3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.05$.

[c3] Difference between this estimate (column 4) using design-based weights and corresponding estimate from respondents (column 5) using final adjusted weights is statistically significant at $p < 0.01$.

[a4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.10$.

[b4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.05$.

[c4] Difference between this estimate (column 5) using final adjusted weights and corresponding estimate from respondents (column 6) using initial adjusted weights is statistically significant at $p < 0.01$.

[1] Illicit Drugs include marijuana/hashish, cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. Illicit Drugs Other than Marijuana include cocaine (including crack), heroin, hallucinogens, inhalants, or prescription-type psychotherapeutics used nonmedically. The estimates for Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine incorporated in these summary estimates do not include data from the methamphetamine items added in 2005 and 2006. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.

[2] Nonmedical use of prescription-type psychotherapeutics includes the nonmedical use of pain relievers, tranquilizers, stimulants, or sedatives and does not include over-the-counter drugs.

[3] Estimates of Nonmedical Use of Psychotherapeutics, Stimulants, and Methamphetamine in the designated rows include data from methamphetamine items added in 2005 and 2006 and are not comparable with estimates presented in NSDUH reports prior to the 2007 national findings report. See Section B.4.8 in Appendix B of the *Results from the 2008 National Survey on Drug Use and Health: National Findings*.

[4] Tobacco Products include cigarettes, smokeless tobacco (i.e., chewing tobacco or snuff), cigars, or pipe tobacco.

[5] Dependence or abuse is based on definitions found in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, NSDUH main study and clinical sample, 2008-2012.